

Can the West contribute more to Human Frontiers?

Japan is urging Western countries to contribute more to the Human Frontier Science Program which supports international research on the brain and molecular biology. But whether they can and should comply is debatable.

When Japan proposed the Human Frontier Science Programme ten years ago with calls for ¥1,000 billion, or several billion dollars, to be invested over 20 years in international research on the brain and advanced robotics, it was greeted with understandable scepticism by both scientists and government officials in the West. They saw this as yet another attempt by Japan to pick the brains of Western scientists to develop new technology.

Since then, however, scientists in both Japan and the West have moulded Frontiers into a small but excellent programme that supports basic research on the brain and biological functions at the molecular level, involving international teams of scientists. The programme, which is based at a foundation in Strasbourg, France, and receives funds from six European nations, the European Union, the United States and Japan, was given glowing reports last year in an external review by an international panel of scientists. The intense and growing competition for the limited number of awards is also testimony to the popularity of the programme. The wider international and interdisciplinary scope of Frontiers research gives it a dimension that national programmes lack. In principle, it deserves to grow.

At present, Japan provides 80 per cent of the annual budget of \$45 million, despite agreement at the last intergovernmental conference in 1992 that other participants together should match Japan's contribution "as soon as feasible". At the same time, Japanese scientists receive only about 7 per cent of the awards. By far the largest share of grants, each worth on average about a quarter of a million dollars a year, go to principal investigators in the United States and European countries.

Japan thus seems to have a legitimate case for asking for larger contributions from the West. At a preparatory meeting later this month it will push for this to be put on the agenda of an intergovern-

mental conference to be held later this year (see page 100).

There are political reasons favouring stronger international support. Japan has been under pressure from Western governments over the past decade to contribute more to basic research. It has responded by putting substantial new funds into domestic government research and into international programmes initiated both in Japan and the West. This week's latest contribution to the European Laboratory for Particle Physics (CERN), which adds to several tens of millions of dollars recently given to CERN by Japan, is one example (see page 102). The CERN funds are a response to intense lobbying by European diplomats in Tokyo and by CERN officials.

Can Western governments expect such contributions to continue if they do not reciprocate? Some might argue that one cannot talk about international research on particle physics in the same breath as international research on the brain. Nevertheless, for officials at Japan's Ministry of Finance, they are much the same thing.

But at this point, proponents of enhanced support hit harsh reality. In Western eyes, the obstacles are significant. Increased funds would come at the expense of national budgets, given current constraints. Furthermore, why put money into this Japanese-inspired programme when there are already excellent national programmes funding research on the brain and molecular biology?

There is no hard evidence that the research is different or better than would have been achieved nationally. But, in high-energy physics the scale of the technical challenges means that progress cannot happen at all without international collaboration. The Frontiers programme has the regrettable problem that internationalism for its own sake will not cut much ice. Promoters of the Frontiers programme must construct a stronger case than has so far been achieved. □

Joining hands in stormy seas

The health of global fish stocks requires new linkages between fishermen and scientists.

There can be few today who are unaware that fishing is an industry in crisis. A combination of modern fishing techniques and a failure to implement safeguards to keep stocks of individual species at sustainable levels means that the availability of many types of fish is declining. Yet this is happening precisely at a time when demand from a rapidly growing population is on the increase.

Caught in this vicious circle, scientists and fishermen find themselves on opposing sides of a frequently acrimonious debate. Researchers find their frustration turning to despair as solid arguments against overfishing prove ineffectual against a powerful industry. Fishermen facing tough economic realities have no incentive to listen to such arguments, however authoritative.

In such a climate, and despite their high credibility in political circles, scientists shouting even louder seem likely to achieve little but more alienation among fishermen. But other approaches are being experimented with (see Briefing, pages 105–109). Off the coast of

Massachusetts, fishermen are being recruited to participate in research projects aimed at understanding the impact of climatic change, one of the explicit goals being to reduce the friction between the two sides. And there is also a growing debate within the scientific community over whether current research strategies, however scientifically valid, remain the most appropriate for the context in which their conclusions are likely to be used.

There are important parallels between the difficulties of linking science advice and fisheries policy and those of making similar links in other areas (such as food safety or environmental carcinogens). Strategies based on the assumption that good science is sufficient to create good policy will prove inadequate. The way forward is to involve as many 'stakeholders' as possible — including scientists where appropriate — in the very task of formulating policy advice. Flexibility on both sides is required to achieve this; the future of the world's fishing stocks surely depends on it. □



Politicians accused of 'shooting from the hip' on human cloning

[WASHINGTON] US scientists and bioethicists reacted with alarm last week to moves by two congressmen to ban human cloning, calling their bills so poorly drafted as to threaten the beneficial research that might otherwise result from the cloning of a sheep by scientists in Scotland.

Brigid Hogan, for example, a professor of cell biology at Vanderbilt University in Nashville, Tennessee, and co-chair of a panel that recommended in 1994 that the National Institutes of Health (NIH) be permitted to fund some human embryo research, expressed a widely held view that politicians were "shooting from the hip".

John Fletcher, a professor of biomedical ethics at the University of Virginia School of Medicine, described the congressional proposals as "profoundly misguided and inappropriate", coming as they do before extensive, reasoned debate has taken place.

There is particular concern that the term

'human cloning' might be interpreted so broadly that a ban would technically cover any research involving the study of the development of human cells in culture after nuclear transplantation from a differentiated adult cell.

The criticism was not only directed at Congress. Others said that the vague wording of an executive ban on federal funding for 'cloning human beings', announced by President Bill Clinton on 4 March, could also stifle exploration of the promising avenues opened by the scientists in Scotland (see *Nature* 385, 810; 1997).

Flanked during the announcement by Harold Varmus, director of NIH, and Donna Shalala, Secretary of Health and Human Services, Clinton also called for a voluntary private-sector moratorium on human cloning until the National Bioethics Advisory Commission (NBAC) considers the issue.

His announcement seemed intended to

respond to some of the initial public outrage about the implications of the Scottish research. In a poll of 519 Americans conducted by *ABC News Nightline* on 24 February — the day after a British newspaper leaked the news of the Scottish work — 87 per cent of respondents said that human cloning should not be allowed, 82 per cent said that it would be morally wrong, and 93 per cent said that they personally would not choose to be cloned (44 per cent called the cloning of animals morally wrong).

But scientists are divided on the evident goal of the president's ban, namely to block the implantation, gestation and birth of cloned human beings. While some call it commendable, others point out that such work is not currently feasible, and indeed is not likely to be done in the future by the kind of people who would pay attention to such a ban. Others argue that such research is blocked by an existing congressional ban on federal funding for human embryo research.

"Is this posturing? The answer is yes," says George Annas, a professor of health law and bioethics at Boston University School of Public Health, adding that the executive ban was a "very, very safe political thing to do".

Two weeks ago, a bill was introduced in the Senate by Christopher Bond (Republican, Missouri) to block federal funds for "research with respect to the cloning of an individual". This defines cloning as "the replication of a human individual by the taking of a cell with genetic material and the cultivation of the cell through the egg, embryo, fetal and newborn stages into a new human individual".

This was soon followed by two bills introduced in the House of Representatives by Vern Ehlers (Republican, Michigan). One would ban federal funding for anyone who "use[s] a human somatic cell for the process of producing a human clone". The other would impose a \$5,000 penalty on anyone attempting such work. Neither bill defines 'human clone'.

Ehlers has already been asked by Newt Gingrich (Republican, Georgia), the speaker of the House, to draw up broad science policy recommendations for Congress (see page 100). At a hearing last week, he portrayed his bills as a pre-emptive move to deflect more drastic congressional efforts to ban all cloning research, including that on animals.

Ehlers argued that "if we don't ban immediately research on cloning humans, we are likely to see a strong movement to ban research on cloning in general". But at the hearing, held by the technology subcommittee of the House science committee, leading

Texan goes gunning for US role in LHC

[WASHINGTON] A Texan congressman has warned the Clinton administration that he will do everything he can to block the United States' planned participation in the Large Hadron Collider (LHC) project at CERN, the European laboratory for particle physics in Switzerland.

Joe Barton (Republican), whose district was to have hosted the Superconducting Super Collider (SSC) abandoned by the United States in 1993, confronted administration officials at a hearing last week on energy department spending plans.

"I'm going to do everything in my power to prevent one dime from being spent on this until we have a written commitment from the members of CERN to participate in a future machine built in the United States," Barton told Martha Krebs, assistant energy secretary, at the hearing. He said that he did not intend "to sit here and allow half a billion dollars to be sent to Europe for this". He added enigmatically that he "knew where the bodies are buried" on the SSC project.

Barton, an industrial engineer and keen SSC supporter even before it was placed in his district, argued that the project would have been almost complete by now if Europe had helped with funding.

Tim Roemer (Democrat, Indiana) echoed Barton's criticism, asking Krebs why the administration had "broken with precedent" by helping to fund construction of the LHC, rather than confining US involvement to the



Halt! Barton (right), seen here with House leader Newt Gingrich, wants European reciprocity.

provision of instrumentation.

Krebs replied that the US contribution would be spent with laboratories and industrial contractors in the United States. The LHC was "the only game in town", she said, and "basically a good deal".

Krebs later said that, in demanding reciprocity from the Europeans, Barton had "raised an issue which is of some concern to us", but which CERN members "have difficulty discussing". She said that she is keeping the door open on the possibility of altering the existing agreement between CERN and the United States before it is finally signed. The deal was initialled in January by herself and CERN's director general Christopher Llewellyn Smith.

A spokesman for Barton said that, even if the appropriations committees accept the administration's proposal for LHC, he will put forward an amendment to kill it on the floor of the House.

Colin Macilwain

scientists responded to the bills with alarm.

Varmus said that Clinton's intention was that any legislation should only be drafted after the NBAC had had the time to think through all the possible applications of cloning. Thomas Murray, a bioethicist at Case Western Reserve University in Cleveland, Ohio, and chairman of the genetics subcommittee of NBAC, pleaded with Ehlers not to take any "irreversible" action on the bills until the commission has had its designated

90 days to study the subject. Ehlers later described Varmus's response as "predictable", and said it would not change his plans.

Varmus also told last week's hearing that ethically defensible cloning of humans might occur in the future, citing its potential use in infertility. Such cloning "if ever to be used would be used incredibly sparingly", he said.

The bills introduced by Ehlers and Bond "will certainly stifle the research on human cells in culture that could give vast benefit to

people", says Roger Pedersen, a developmental geneticist at the University of California, San Francisco.

He says that Ehlers' two bills would be "disastrous" because of their use of the term 'human clone' which, technically, applies to any cell population arising from a single ancestral human cell. In principle, the bills could ban the study of how the cytoplasm of the human egg induces differentiated adult cells to return to totipotency, the state in which a cell may give rise to any kind of tissue.

They could also block studies of cellular senescence (ageing), as well as efforts to culture embryonic stem cells to provide tissues for auto-transplant.

Ehlers argues that Pedersen is trying to put his bills in "the worst possible light", and adds that he does not think that they would be interpreted as narrowly as Pedersen suggests. "I'm not trying to limit science," he said, but "most" of the experiments described "could easily be done on other mammals".

Some researchers fear that the Bond bill could also be so strictly interpreted as to stifle research. They point, for instance, to the possibly wide application of its language forbidding "research with respect to the cloning of an individual".

Experts say these ambiguities make clear that the bills will have to be substantially and precisely rewritten before they come to votes in the House and Senate. That they will ban the birth of cloned human beings seems all but certain, given the political climate. But, at stake in the revision process will be "whether we're going to permit research at the [human] embryo level with the clear prohibition that it could never be implanted in an animal or a human", says Annas.

Some argue that the Clinton ban is already redundant, as Congress banned federal funding for human embryo research in its 1996 and 1997 budgets, expanding a 1994 Clinton ban on federal funding for the creation of human embryos for research purposes to include research on spare embryos left over from *in vitro* fertilization.

The congressional ban, attached as a legislative rider to bills funding the Department of Health and Human Services, forbids federal funding of research in which human embryos are "destroyed, discarded or knowingly subjected to risk of injury or death". It defines embryos as being "derived by fertilization, parthenogenesis, cloning, or any other means from one or more human gametes".

Clinton said last week that this ban — which he tried to reverse in his budget proposals in 1997 and 1998, on the grounds that it does not belong in legislation — does not "explicitly cover human embryos created for implantation". Nor does it cover all government agencies.

Meredith Wadman

UK embryo research law 'may need changes'

[LONDON] British legislation may have to be amended to enable the government to prevent techniques recently used to clone an adult sheep from being applied to human cells without controls, according to the head of the organization responsible for licensing research on human embryos.

But Ruth Deech, the chair of the Human Fertilisation and Embryo Authority (HFEA), warned a parliamentary committee last week that rigid legislation, in particular an outright ban on human cloning, could stifle potentially beneficial research.

Instead, she said, the authority would prefer to see an extension of its existing powers, giving herself or the Secretary of State for Health the authority to judge whether a particular line of research should be permitted. "We feel that the act may only require some tweaking to deal with public concern."

According to Deech, who was addressing a hearing of the House of Commons Select Committee on Science and Technology, whether new legislation is needed depends on the precise meaning of the word 'embryo' contained in the legislation passed in 1990 that set up the HFEA.

Uncertainty has arisen over whether the term 'embryo' as used in the act is sufficiently broad to cover the techniques developed by Ian Wilmut and his colleagues at the Roslin Institute in Scotland (see

Nature 385, 757; 1997). The researchers took a nucleus from a developed udder cell and inserted it into an embryo cell whose own nucleus had been removed.

Wilmut himself told the House of Commons Select Committee on Science and Technology last week that, in principle, he expected it would be possible to use a similar technique to clone humans "within two years".

It is this prospect that has forced the HFEA to consider whether it has the authority to regulate such experiments under its current terms of reference. "We are 100 per cent convinced that we are able to deal with everything except the one technology that the Roslin Institute applied," Deech told the committee.

Graham Miles, the HFEA's chief legal adviser, said that discussions are taking place between the HFEA and the Department of Health about whether the terms embryo, gametes and fertilization are sufficiently broad to allow the authority to regulate the type of cell-fusion experiments conducted at Roslin.

If the answer is negative, said Deech, it would be necessary to ask parliament to revise the wording of the act and to broaden the definition of the word 'embryo' accordingly.

Such a move would inevitably offer the opportunity to individual members of parliament opposed to embryo research in general to seek a total ban on all activities relating to human cloning. David Alton



Deech: favours flexibility.

(Liberal Democrat, Mossley Hill) has already called for the suspension of all cloning research in animals and humans.

Alton supports calls for an international ban. "We don't know where the DNA highway will lead, and until there has been proper discussion it is legitimate [to stop and wait]," he says.

But Deech, who claimed that Britain has "led the way in the regulation of *in vitro* fertilization and allied treatment", warned that a blanket ban could have serious drawbacks. "If the act does have to be amended, we should maintain a flexible approach and leave the door open to the potential benefits of this technique," she said.

Deech's call for caution in amending the act was echoed by George Radda, secretary of the Medical Research Council. He told the select committee: "If the procedures at Roslin Institute lead to a fundamental understanding of how to turn on genes in an adult organ, that has enormous possibilities for regenerating a failed heart or a damaged brain. I want us to be able to support the best science we can that is going to advance medicine in the best way."

David Dickson

ESA pulls back from 'fair shares' funding

[MUNICH] European space ministers have agreed partly to abandon their commitment to the principle of *juste retour*, the controversial mechanism by which countries receive industrial contracts for projects funded through the European Space Agency (ESA) in proportion to their subscription to the agency.

The decision was made at a special ministerial meeting in Paris last week, and represents a victory for those who have been arguing that the award of contracts should be based more heavily on technical competence and value for money.

Juste retour was introduced into ESA's convention when the agency was set up in 1975 with the aim of ensuring that each member state receives adequate support through ESA for its national aerospace industry. But the system has come under increasing strain over the past few years, as the allocation of contracts on a 'fair shares' basis inevitably drives up costs and reduces competition.

The system is also complicated to run, and several countries have not, in fact, received their 'fair shares'. Italy, ESA's third largest contributor, has seen its industrial return fall in the past few years well below the minimum guaranteed level of 0.96, approximately equivalent to 96 per cent of its contribution to ESA. In contrast, Germany and France have return coefficients well over 1.1, a reflection of both better organized aerospace lobbies and stronger industries.

Some of the smaller countries, particularly Sweden and Switzerland, have fared even worse than Italy. Their industries are too small to act as prime contractors, and they are dependent on prime contractors from larger countries subcontracting to them on a basis that takes account of geographical considerations. ESA has not supervised this process stringently enough.

After more than a year of debate, Europe's space ministers, meeting as ESA's council of ministers, have agreed that the rules of *juste retour* should be relaxed after 2000. Meanwhile, the deficits in industrial contracts accrued by Italy, Sweden and Switzerland will be corrected, and the slate wiped clean.

ESA's science programme, the agency's only mandatory programme, will not abandon *juste retour* entirely, but will operate it with more flexibility. The guaranteed minimum return coefficient will be lowered from 0.96 to 0.9, and the books will be balanced every three years, rather than accumulating indefinitely.

Optional ESA programmes, such as the launcher Ariane and participation in the US space station, in which participation by member states is voluntary, will make a more radical move. Industrial contracts will be awarded

solely on the basis of value for money, and the contributions of participating member states to ESA will be adjusted according to their success in winning contracts.

The procedures for contract procurement will also be revised in the next three years to increase fairness — for example by including potential subcontractors as well as prime contractors — and efficiency, and to protect the interests of small and medium-sized enterprises involved in developing space technologies.

Ian Taylor, Britain's science and technology minister, welcomed what he called a "fundamental change of attitude" in ESA. "Negotiations have been tough," he says. "But member countries have come to realize that the old fixed expectations of guaranteed work and contracts will not do for today's fast-moving world markets." Germany has also welcomed the changes as a chance to

increase competitiveness.

Sergio De Julio, president of the Italian space agency ASI, says that Italy is now satisfied with ESA's plans to normalize Italy's industrial return by 2000. This will be achieved primarily by guaranteeing that Italian industry will receive major contracts to develop parts of ESA's contribution to the space station. At present, Italy's industrial return deficit is well over ECU100 million (US\$88 million).

ESA's council of ministers will monitor the response to the new regulations over the next few years and will consider the possibility of entirely abandoning the geographical considerations in awarding contracts after 2000. But this is unlikely to happen without a fight. Open bidding has many opponents among smaller countries whose industries may not be able to survive in a highly competitive environment.

Alison Abbott

Mars risks are low but not zero, says NRC

[WASHINGTON] A potential threat to plans for a spacecraft mission to bring back samples from Mars has been removed by a report which concludes that Mars landers will not necessarily need to undergo an expensive sterilization process.

But the report, published last week by the US National Research Council (NRC), says that rock samples returned to Earth should still be quarantined until proved harmless, as the risk of an adverse biological effect, while low, "is not zero".

Past planetary missions have been designed to protect against "forward contamination" — the risk of introducing terrestrial microbes to another planet. The Viking landers of the 1970s underwent an elaborate sterilization process in which the entire spacecraft was heated before launch.

Managers at the National Aeronautics and Space Administration (NASA) had been concerned that a Mars mission planned for 2005 would be held to the same standard, as recommended by a 1992 NRC report. But the new report says that other techniques, including radiation and chemical treatment, "may be more effective and less costly".

Kenneth Nealson, a microbiologist at the University of Wisconsin at Milwaukee, who chaired the study, thinks the agency "will be delighted with these recommendations", as it should help the mission to stay within its relatively modest budget. NASA plans to increase its \$100-million annual budget for Mars exploration by about \$50 million a year to make possible a launch in late 2004, with arrival on Mars in 2007 and samples returning to Earth in 2008.

The report makes little mention of last

Artist's conception of sample return mission.

summer's controversial claim of evidence of fossil life in a martian meteorite. But it says that there are other reasons to suspect Mars may have harboured life in the past, and that it may yet exist beneath the surface or near hydrothermal vents.

Even though "contamination of Earth by putative martian microorganisms is unlikely to pose a risk of significant ecological impact or other harmful effects", the report recommends that no uncontained samples be returned to Earth. Apart from being prudent, such a strategy would help NASA to avoid legal challenges to a sample-return mission, says Nealson.

The panel also advises NASA to establish a research facility for handling returned samples at least two years before the mission is launched, so that a team of scientists can practise working with samples before the real ones arrive. That would help to prevent the mistakes that were made when lunar samples were returned by Apollo astronauts, and quarantine was breached on several occasions.

Tony Reichhardt

NASA

Japan bids to widen brain research funds

[TOKYO] The Japanese government is to renew attempts to persuade other leading industrialized countries to increase their financial support for the Human Frontier Science Program (HFSP), the international research programme in molecular biology and brain science that was launched on Japan's initiative in 1987.

The issue of future funding will be a key item on the agenda of a meeting at the HFSP headquarters in Strasbourg on 24–25 March. The meeting has been called to agree on items to be discussed at an intergovernmental conference which is expected to take place in the United States later this year, perhaps to coincide with the G7 summit in Denver.

At the Strasbourg meeting — the first intergovernmental meeting on HFSP to be held for five years — Japanese officials hope to discuss the programme's funding 'inequality'; despite the success of the programme, Japan still pays for 80 per cent of its

annual budget. The United States increased its contribution to the programme by \$500,000 to \$4 million last year. But most Western countries in the programme appear reluctant to increase their financial support.

The meeting comes at a critical time. The award rate for HFSP grants has fallen to 10 per cent, making it difficult for even highly regarded applications to win funding. HFSP officials are concerned that this may discourage good researchers from investing the time necessary to make applications in the future.

HFSP was started on the initiative of Yasuhiro Nakasone, the then Japanese prime minister, to pursue developments in the life sciences through international cooperation. At present, the United States and Canada jointly provide roughly 10 per cent of the programme's \$45 million annual budget. France, Germany, Italy, Switzerland, the United Kingdom and the European Commission together contribute a further 10 per cent.

Principal investigators from the participating countries, who head multi-disciplinary international research groups, are entitled to compete equally for awards, regardless of the amount their country contributes.

At the previous intergovernmental conference, held in Tokyo in 1992, agreement was reached in principle on the desirability of achieving an equal match between the contributions of Japan and those of the other participating countries. But there has been only a slight increase in the share from other countries in the past five years.

HFSP grants are worth on average \$240,000 a year for up to three years, and 45 such grants were awarded in 1996. Of these, 38 per cent went to principal investigators in the United States, 15 per cent to Germany, 11 per cent to the United Kingdom, 13 per cent to France and 7 per cent to Japan.

In addition to the question of future funding, an agenda for the intergovernmental conference planned for later this year including possible new members, a change in the scope and activities of the programme, as well as various management issues, is expected to be finalized.

The meeting follows a review of the HFSP carried out by external consultants and published last year. This described the initiative as an "excellent and valuable programme", some of whose aspects are unique among science programmes, for example its encouragement of research collaboration at global level.

One participant is Terrence Sejnowski, a computational neurobiologist at the Salk Institute of Biological Studies in San Diego, California, who helped to organize a consortium to study the function and mechanisms underlying sleep with a grant from the programme. Sejnowski says that the programme has been "extraordinarily successful" in making possible research advances and publications.

But the HFSP continues to face the major hurdle that many of the participating countries consider that they can spend research funds more effectively through their national research organizations rather than through an international secretariat. Indeed, one British researcher, who previously won an award but has been unsuccessful in three recent applications, says he finds it easier to secure funds from funding organizations in the United Kingdom, despite its financial difficulties, and would oppose money being taken from the domestic budget and given to the HFSP.

After the meeting in Strasbourg, the programme's board of trustees will select the recipients of this year's grant and fellowship awards.

Richard Nathan

Task force will review US science policy

[WASHINGTON] A bipartisan task force is to be set up in the House of Representatives to conduct a wide-ranging review of all aspects of US science policy.

The review, due for completion by the end of next year, will be led by Vernon Ehlers (Republican, Michigan), the vice-chair of the Science committee and a former nuclear physicist. It has been set up after discussions with Newt Gingrich (Republican, Georgia), the speaker of the House, and James Sensenbrenner (Republican, Wisconsin), the chair of the Science committee.

Ehlers said that both Gingrich and Sensenbrenner had "expressed an interest in a complete review of national science policy". According to congressional staff, Ehlers has obtained approval to pay for the salary of a staff director responsible for implementing the review. The review team is expected to hold hearings in the Congress, and to issue a report within the life of this Congress, which ends in December 1998.

Ehlers outlined his thinking during a meeting last Thursday (6 March) with the President's Council of Advisors on Science and Technology (PCAST). "We have to rethink the science policy of the nation," Ehlers told PCAST. "We don't have a science policy; we have a budget policy, and I want to correct that."

He pointed out that there had been several recent efforts by non-governmental groups — including a panel of the National Academy of Sciences chaired by its former president, Frank Press — to update the work of Vannevar Bush, whose 1945 report



Ehlers: a physicist turned legislator.

Science: The Endless Frontier laid the foundations for US science policy after the Second World War. He welcomed these efforts but said that, "unless Congress and the administration buy into the process, it won't get anywhere".

Ehlers is a nuclear physicist who studied and lectured at the University of California at Berkeley before running the physics department at Calvin College, Grand Rapids, Michigan. He is particularly interested in undergraduate science education. But the scope of the review is expected to be far broader.

A similar review was undertaken in the mid-1980s by the then House science, space and technology committee, but it made slow progress and its staff was disbanded before a final report had been completed.

Veterans of that review suggest that the Ehlers initiative may fare better. They argue that restricted budgets mean that both government agencies and congressional committees are interested in receiving new guidance on setting priorities, which should lend impetus to the review process.

Together with the recent formation of a Science and Technology Caucus in the Senate, the review has been welcomed by science lobbyists in Washington, who see it as evidence of a resurgence of bipartisan support for science and technology in the Congress. **Colin Macilwain**

BSE surveillance and safety measures still found wanting

[PARIS] One year after the United Kingdom announced a possible link between bovine spongiform encephalopathy (BSE) and Creutzfeldt-Jakob disease, many member states of the European Union have still not taken adequate steps to detect BSE or to prevent a further spread of the disease.

This is the conclusion of an internal European Commission document, *Risk factors and surveillance with regard to BSE in the community*, based on short inspections in 13 member states last autumn (the United Kingdom and Portugal were evaluated separately).

The report says the quality of surveillance systems for detecting BSE cases varies widely, while the fact that farmers are not always compensated in full for declaring cases acts as a disincentive. It calls for laboratory examination of all suspected BSE cases, complete record-keeping, and better training. The commission should establish a reference laboratory to ensure uniform application of standardized procedures, it adds.

The number of expected cases of BSE in continental Europe has been estimated at several thousand — although only a few hundred cases have been officially reported (see *Nature* 382, 4; 1996). This does not mean that an epidemic is going unreported, emphasizes a commission official but, rather, that there is a need to have good systems in place as a preventive measure.

Of more immediate concern is the risk that even low levels of BSE could increase because of insufficient controls on the animal feed in which the disease is transmitted. The European Commission has introduced a ban on feeding mammalian protein to ruminants, but its inspectors found that feed labelled as containing no meat or bone meal often contained as much as 5 per cent.

The most likely explanation is cross-contamination — most feed mills use the same equipment to produce all feed, so meal destined for pig and poultry feed can find its way into cattle feed. One commission official says that the only solution is either to have separate mills for pig and poultry rations and for ruminant feed — as is happening in Ireland — or to ban mammalian protein from the feed of all farm animals, a step taken belatedly in the United Kingdom last March.

Most member states say they are implementing a European Commission directive stipulating that rendering plants must convert to batch-processing at 130 °C and 3 bars pressure for at least 20 minutes. But many are behind in doing so, while France insists its existing arrangements are adequate. Inspectors found evidence of non-compliance with procedures in several states.

Declan Butler

'Poor security' blamed for loss of Russian fossils

[MOSCOW] A meeting of the biology section of the Russian Academy of Sciences was told last week that the disappearance of specimens from the academy's botanical, zoological and palaeontological collections has been largely the result of lack of funds for proper security measures.

The meeting was held to respond to allegations that dinosaur skulls missing from the academy's Institute of Palaeontology in Moscow may have been smuggled out of the country (see *Nature* 384, 499; 1996).

It discussed a resolution drawn up by three corresponding members of the academy which stated that, although the institution's collections were of substantial scientific and cultural value, insufficient funds for security had resulted in numerous thefts.

But the meeting failed to adopt the resolution, which attempted to respond to the smuggling charges by criticizing foreign researchers for holding on to specimens from the institute's collections for longer than had been agreed.

The panel confirmed that the palaeontology institute has lost 23 amphibian skulls, two mammoth tusks and an ammonite collection, in addition to three dinosaur skulls and two jaws which disappeared last year.

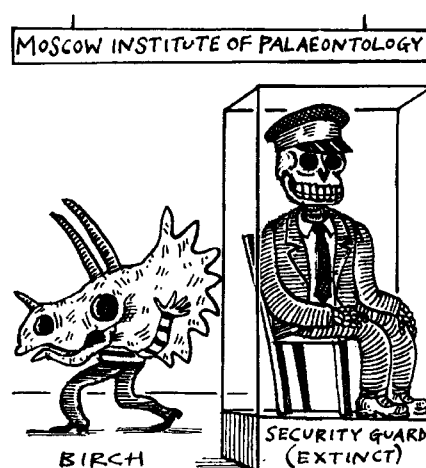
But the institute is not alone in experiencing losses. Between 1992 and 1996, for example, the Zoological Institute in St Petersburg lost articles made of mammoth tusk, three rhinoceros horns and one mammoth tooth, as well as 13 species of tropical beetles. Rare wood and tobacco specimens have disappeared from a botanical museum as well as a sculpture weighing more than 120 kg.

The authors of the draft resolution blame what they describe as "the imperfection" of Russian law which "prevents control of the biological objects market, and attracts to it criminal elements". They also claim that a "special problem" is the refusal of some Western specialists to return items loaned to them — naming in particular scientists at the University of Bristol in the United Kingdom.

But one of the UK researchers concerned dismisses such charges, pointing out that the Bristol group had been informed by the institute that there would be "no problem" in holding on to the specimens because one of the researchers is a Russian national who is being supported by the Royal Society.

Aleksei Rozanov, director of the Institute of Palaeontology, and his deputy, Igor Novikov, claimed that accusations that staff from the institute may have been involved in the disappearance of specimens was "unfair".

But Vladimir Sokolov, chairman of the academy's biology section, who chaired the



meeting, asked Novikov, "how could it happen that an internationally known scientific body did not protect its reputation?". On the question of loans to Western institutions, he said that he was unable to believe the situation as it had been reported.

Novikov said that the failure to return specimens on time is a worldwide practice, admitting that Russian scientists sometimes keep foreign material for years longer than agreed. He also reported that one of the two dinosaur jaws that he had earlier declared to be missing has since been found in the museum.

Rozanov told the meeting that the institute had lost hope that either the national or international police forces could help, and that this was why neither had been contacted about the latest thefts. "Instead we made direct contact with customs authorities to try to prevent the fossils leaving Russia illegally."

According to Rozanov, the power of the black market in this field is such that he was not willing to name the institute official who was in touch with the customs authorities for fear that he might be killed. He said that shortage of money meant that unofficial contacts were the only way open to the institute to address the problem.

Only after the first thefts had money been found to provide museum showcases with locks, and a basic alarm system had not been installed until 1995. (The equally valuable museum of the St Petersburg Zoological Institute still has no alarm.) A sum of US\$167,000 is needed to ensure the safety of the collections in the Institute of Palaeontology.

Alexei Yablokov, head of the Centre for Russian Environment Policy, whose call for steps to protect the collections had prompted the meeting, said he believed that further measures were needed.

Carl Levitt



Kwun: will steer new science legislation

Physicist takes top Korean science post

[SEOUL] South Korea has replaced its science minister for the third time in eight months. Last week, President Young-Sam Kim appointed Sook-il Kwun, a professor of physics and former dean of science at Seoul National University, to head the Ministry of Science and Technology.

Senior science advisers to the Korean government expressed relief that the ministry is once again in the hands of a scientist with a sound understanding of the importance of basic research. One of Kwun's key tasks will be to push forward new legislation for the promotion of science and technology.

The proposed law, already in draft form, is modelled partly on the basic law for science and technology passed recently in Japan. But, with presidential elections due next February, Kwun will have less than a year in office to see the law passed.

Kwun's appointment is part of a cabinet shake-up following a series of scandals that has rocked Kim's government. Kwun replaces Yong-jin Kim, who served barely two months before getting embroiled in a scandal over his role in extending bank loans to the debt-ridden Hanbo business group in his former position as head of the Office of Bank Supervision.

Yong-jin Kim had succeeded Pon-young Koo, an economist who was promoted to become the senior representative of South Korea at the Organization of Economic Cooperation and Development only a few months after he had replaced Kun-Mo Chung as science minister last August. Chung was serving his second term as science minister and had been in office for more than a year — a comparatively long period of office in South Korea these days. He is one of the founding fathers of South Korea's first graduate university, the Korea Advanced Institute of Science and Technology (see *Nature* 364, 379; 1993).

David Swinbanks

California agency charged with delaying toxin reports

[SAN FRANCISCO] California's environmental research agency was accused last week of diluting and delaying scientific reports that may have led to stricter regulations on potentially dangerous chemicals. The accusations were made by university scientists, former employees and environmental activists in testimony before the state legislature.

Their concern reflects fears expressed by legislators when the agency was established in 1991 about the dangers of shifting responsibility for public health and scientific risk assessment from the Department of Health Services to the California Environmental Protection Agency. The legislators were worried that political appointees might interfere with the scientific operations of the office to justify particular policy initiatives.

Last week's hearing was called by Senator Byron Sher (Democrat, Palo Alto) in response to complaints that the Office of Environmental Health Hazard Assessment (OEHHHA) had inappropriately interfered with and altered public risk assessments of environmental toxins.

The office was also accused of quashing various studies, including one on airborne pesticides, and allowing its budget to decline to a precarious level. "We thought it was important to have a hearing and give the director a chance to respond," Sher said.

At the hearing before the Senate Committee on Environmental Quality, university scientists who review OEHHHA's work complained that reports on environmental tobacco smoke, lead and vehicle emissions had been watered down by state employees and held up for several years.

Representatives of the Natural Resources

Defense Council, an environmental group, identified 70 chemicals known to cause birth defects or cancer about which it claimed the state had failed to warn the public.

The senators also heard allegations that OEHHHA was cutting back on staff to a crippling level and that it had terminated several important studies, including one on the impact of pesticide drift on the number of hospital admissions of patients with asthma in Lompoc, California.

Richard Becker, director of the OEHHHA, said in response that he was committed to sound scientific analysis, promptly delivered, and that tensions in his office were the result of his demands for greater openness, greater objectivity and a closer look at all possible mechanisms and modes of action of particular chemicals.

He said the state had already issued warnings on some 583 toxic chemicals, and would continue to carry out its mandate to warn the public about dangerous chemicals.

But Sher said after the hearing that he was convinced that the problems at the agency were real, and expressed particular concern about pressure from outside interests — such as lead industry representatives — who may have contributed to the three-year delay on the lead report. He asked Becker for information to explain the cause of the hold-up.

The report on lead's health risks to children and adults was released last October after five years of preparation; the report on environmental tobacco smoke, which was issued two weeks ago, took six years to complete; and a study on the health risks of diesel exhaust has been in preparation for more than four years.

Sally Lehrman

CERN raises funds for antimatter detector

[MUNICH] Drawing on substantial support from Japan, CERN, the European Laboratory for Particle Physics, is to build an experimental facility, the Antiproton Decelerator, to study the characteristics of antihydrogen and other exotic atoms.

The decelerator will cost SFr7 million (US\$4.7 million). Japan is contributing at least half of the costs of building the machine, and will also spend a further SFr10 million on a major antimatter research project to be carried out using it. The facility will be built by transforming an existing machine at CERN, the Antihydrogen Collector, which collected the first antihydrogen atoms ever detected, produced by the now-defunct LEAR (low-energy antiproton ring) facility in 1995.

Budgetary restrictions had threatened the development of this pioneering work, which was repeated more recently at Fermilab in the United States. But CERN announced this week that a fund collected from countries including Denmark, Germany, Italy, Japan, Poland and the United States, will allow physicists to explore whether a world made entirely of antimatter would have exactly the same characteristics as the world of matter.

The opportunity will allow CERN to carry out particle physics experiments in the five-year period after its major accelerator, the Large Electron-Positron collider, is switched off and before its successor, the Large Hadron Collider, comes into operation.

Alison Abbott

Economic impact of MIT spin-offs larger than Thailand's GDP

[BOSTON] The substantial economic impact of the Massachusetts Institute of Technology (MIT) has been confirmed in a report which calculates that there are 4,000 active companies that were either founded by MIT graduates, faculty or staff, or were spun off from MIT laboratories.

Worldwide, these companies employed 1.1 million people in 1994, and generated \$232 million in sales. In the United States alone, MIT-related companies employ 733,000 people at more than 8,500 plants and offices, and account for one out of 170 jobs in the country.

Included are many top corporations, such as Campbell Soup, Digital Equipment Corporation, Gillette, Hewlett-Packard, Intel, McDonnell Douglas, Rockwell International, Raytheon and Texas Instruments.

The report, published last week, says that, if these corporations were to form an independent nation, their combined revenues would make it "the 24th largest economy in the world", with a gross domestic product between that of Thailand and South Africa.

The report is the first to assess the importance of an individual research university to the national economy, and it underlines the value of federal government support. "About 90 per cent of these companies have been founded in the past 50 years, in the period of the great research partnership between the federal government and research universities," says MIT's president, Charles Vest.

"The development of these business enterprises is one of the many beneficial spin-offs of federally funded research which has brought great advances in such fields as healthcare, computing and communications." Each year, about \$370 million is spent on research on the MIT campus, \$270 million of which comes from the federal government. Another \$338 million goes to research at the MIT-affiliated Lincoln Laboratories, primarily from the Department of Defense.

The study's director, Wayne Ayers, chief economist for BankBoston, says the study tried to explain how MIT has been so effective in generating new businesses in a "national economy that is increasingly emphasizing innovation". Factors cited by graduates who went on to start companies include encouragement from faculty mentors, access to the latest technology and an "entrepreneurial spirit" instilled by the university.

"MIT is not the only university that has had a national impact of this kind," says Ayers. But he says it is impossible to compare MIT's contribution to that achieved by other major research institutions until similar analyses are made.

Steve Nadis

US seeks greenhouse gas cuts from the Third World

[LONDON] US delegates to the United Nations (UN) climate convention clashed with those from developing countries, the European Union (EU) and environmentalist groups last week over US proposals to include commitments from the developing world in a proposed protocol for reducing greenhouse gas emissions.

At present, only developed countries are required to reduce their greenhouse gas emissions, according to an agreement on the terms of a future protocol reached two years ago. The protocol itself has to be agreed at the next annual conference of the UN climate convention, which is due to be held in December in Kyoto, Japan.

But at a meeting of signatories to the convention in Bonn last week, at which countries presented their initial negotiating positions, the US delegation — apparently with one eye on potential difficulties in Congress — suggested that the original terms of the proposed protocol may need rewriting.

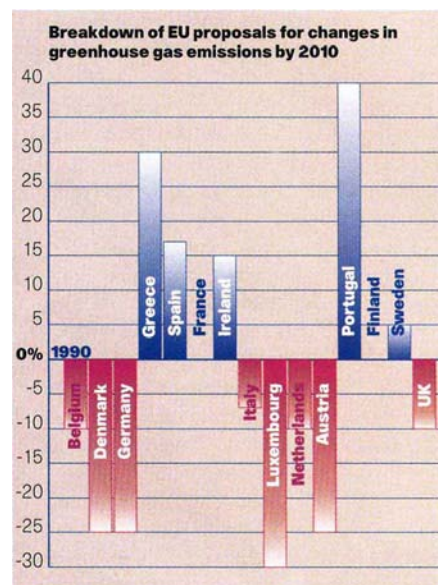
US officials have tabled suggestions that some developing countries should be encouraged to begin reducing their emissions voluntarily, with all countries taking some steps by 2005 — possibly according to a formula in which wealthier countries would make proportionately higher reductions, termed differentiation.

China, which leads the block of 77 developing countries known as the G77, has threatened to pull out of the talks unless the US proposal is withdrawn. Privately, however, some G77 delegates are understood to be attracted to the idea of differentiation between responsibilities as a possible basis for calculating commitments after Kyoto.

The EU, the Alliance of Small Island States (AoSIS) and environmentalist groups, such as the Climate Action Network, have also criticized the plans. They prefer the simpler, but politically more challenging, goal of a standard reduction in emissions for all developed countries only. They also want to avoid complicating talks further by ensuring that commitments from developing countries are not discussed until after Kyoto.

AoSIS and environmentalist groups favour the toughest measure: a 20 per cent cut in greenhouse emissions by 2005. EU environment ministers announced a target of reducing emissions by 15 per cent of 1990 levels by 2010. A 10 per cent reduction would be achieved through measures in individual countries (see figure, above right), with a further 5 per cent reduction through other policies and measures.

Australia, a large coal producer, and the oil-exporting countries are calling for com-



pensation for any revenues lost as a result of the protocol, which may force their customers to switch to 'greener' forms of energy.

One environmentalist group, the London-based Global Commons Institute, has emerged as a surprise supporter of the US plans which, it believes, "have the potential" for providing a more equitable basis of emissions reductions.

The institute's director, Aubrey Meyer, points out that the European position is also based on a form of differentiation, as poorer countries, such as Greece and Portugal, have been allowed to increase their emissions. He believes the US plan could "go far with one or two big G77 names" on its side.

But the prospect of reopening the terms of the protocol upsets campaigners such as Merylyn McKenzie Hedger of the World Wide Fund for Nature. "[This idea] risks stalling everything," she says. "Let's concentrate on what's achievable." Nevertheless she believes that a split in G77 ranks could increase support for a flat-rate reduction.

But one developed country delegate says the United States is unlikely to risk derailing the protocol. Rather, he believes the US plans are likely to help shape the terms of a future protocol that includes developing countries, but which will be settled after the Kyoto talks.

The US plan also supports the creation of a market in greenhouse emissions. Countries needing to make small reductions to meet a national target could agree to an extra reduction in exchange for cash or environmental technologies. A country purchasing this 'entitlement' could credit it towards its own emissions target. Another idea is to 'borrow' emissions from a future period to assist current compliance.

Ehsan Masood

Max Planck institutes face broad cuts, but two escape closure

[MUNICH] About one-fifth of the Max Planck Society's research institutes in west Germany face staff cuts in the next few years, according to the society's president, Hubert Markl, speaking after a senate meeting last week. But two new institutes will be established in east Germany, enabling the MPS to reach its target of having 20 per cent of its research strength in the east.

The society's senate agreed the details of staff reductions required to meet government demands to reduce staff numbers by 512 by 2000 (see *Nature* 383, 566; 1996). Only two institutes will be closed completely — the Institute for Biology in Tübingen and the Institute for Behavioural Physiology in Seewiesen, Bavaria — rather than four as initially proposed.

The Institute for Aeronomy in Lindau, Lower Saxony, will have its staff numbers and budget halved over the next decade. But it will survive as a smaller institute, with two departments focusing on planetary and solar physics. The Institute of History in Göttingen will be saved thanks to the social sciences section of the MPS, which has found a way of sharing the required 50 job losses across several institutes.

Fisheries row looms at North Sea meeting

[LONDON] A conflict of interest between European Union fisheries and environment ministries is threatening to neutralize a special joint ministerial meeting in Bergen, Norway, this week to discuss the environmental implications of overfishing in the North Sea. Specific wording on measures, targets and timetables to tackle overfishing is understood to have been removed from the draft conclusions.

Environment ministers are understood to want prompt action to regulate fishing gear, to reduce the quantity of unwanted species caught, and to designate 'protected areas' where fishing is banned. But fisheries departments and the European Commission, are known to favour including such measures within an expanded Common Fisheries Policy, which is due for review in 2002. (See Briefing, pages 105–110).

Industry excels in life sciences citations

[LONDON] Publications resulting from life sciences research sponsored by private industry are more highly cited than UK research as a whole in the field, according to a report published last week by the Science Policy Research Unit at the University of

Sussex. The report, *The changing shape of British industrial research*, contradicts claims that the commercial sponsorship of research stifles open publication. Indeed, it notes that some British companies publish more scientific papers than many universities.

An analysis of papers with at least one author from industry published between 1981 and 1994 found a significant increase in publications in the pharmaceutical and biotechnology fields, but a reduction in the electricity and electronics sectors. That partly reflects reduced government spending on military research and development.

Minister admits he misled over Gulf War research

[LONDON] A British defence minister has admitted to misleading parliament while giving answers to questions about 'Gulf War syndrome'. Earl Howe, a junior defence minister, was asked last year about the cause of death of animals found in the Saudi and Kuwaiti deserts at the end of the Gulf War.

Howe replied that, following an examination of carcasses at the Centre for Tropical Veterinary Medicine at the University of Edinburgh, it had been concluded that the animals had died of natural causes. But internal defence ministry enquiries have since revealed that samples were never sent to Edinburgh. And, in a statement to the House of Lords last week, the minister revealed that no tests had been carried out at any research establishment.

'Better security needed' to protect medical data

[WASHINGTON] The United States needs a new set of standards and regulations, as well as more consumer pressure, to remedy an electronic system in which personal medical data is widely available and inadequately safeguarded, the National Research Council reported last week.

The panel's recommendations include the encryption of information transmitted on the Internet, electronic 'firewalls' denying access to unauthorized users of electronic health records, and sanctions for security breaches by authorized users, such as leaving patient information on an unattended computer screen. But critics say that the panel did not go far enough, partly because it did not challenge the idea that widely available electronic health data must be an integral part of an effective and cost-efficient system.

Italian nominated to lead Europe's space agency

[ROME] The European Space Agency is to have its first Italian director general. A special meeting of the agency's council of ministers

last week agreed to nominate Antonio Rodotà, the 52-year-old head of the Rome-based aerospace company Alenia Spazio, to succeed Jean-Marie Luton when his term of office expires at the end of 1997.

The nomination is expected to be approved at a full meeting of the council next week. Rodotà is an engineer by training and has extensive experience in management within the space industry, where he has a reputation for efficiency. His nomination follows agreement among the agency's four largest member states to support an Italian candidate for director general, a British candidate for the technical services directorate, a German candidate for industrial policy and a French candidate for administrative services. All four posts fall vacant this year.

Cancer researcher is cleared of misconduct

[WASHINGTON] Bernard Fisher, a cancer researcher at the University of Pittsburgh in Pennsylvania, has been told by the Department of Health and Human Services' Office of Research Integrity (ORI) that he has been exonerated of charges of scientific misconduct brought by the government in 1994. The charges alleged fraud in a major federally funded breast cancer study.

Fisher came under fire when he headed the National Surgical Adjuvant Breast and Bowel Project, which was seeking the best treatment for breast cancer. A collaborator who admitted accepting women not meeting the study's inclusion criteria was found guilty of misconduct by the ORI in 1993. The National Cancer Institute dismissed Fisher after media coverage of the case in 1994. Fisher is suing the government for damage to his reputation.

Brookhaven lab director announces resignation

[WASHINGTON] Nick Samios has resigned as director of Brookhaven National Laboratory, currently at the centre of controversy following the discovery that tritium was leaking from a research reactor fuel storage vessel (see *Nature* 386, 3; 1997). Samios, 64, announced last week that he will leave the director's post on 30 April. He said that he had informed the laboratory's management contractor, Associated Universities Inc, a year ago of his intention to leave this spring, after 15 years in the post.

Critics of the lab, on Long Island, New York, have been calling for his departure since the leak was reported in January. But Samios' resignation was greeted with dismay at Brookhaven, where his energetic leadership style is credited with the scientific revival of the laboratory. Samios says he will return to research at the laboratory.

Fisheries science: all at sea when it comes to politics?

Despite continued warnings by scientists of the disastrous impact of fishing practices, stocks continue to disappear. While some say that scientists should shout more loudly, others suggest it is time for them to change their tactics.

World fish stocks are in crisis. Almost two-thirds of marine stocks in the Pacific and Atlantic oceans are being fully exploited or have already been overfished, even if some are recovering slowly. Future projections predict a steadily widening gap between the world's demand for fish and the ability of the oceans to meet it.

Many scientists claim that the situation is the inevitable result of the failure of governments to heed their warnings of the dangers of overfishing; and where they have listened to this message, failing to adopt and forcefully police policies designed to prevent this.

This is particularly so in the developed world, where the wishes of powerful fishing industry lobbies to retain higher catch quotas can frequently outweigh the advice of scientists who suggest a reduction in order for fish to be sustainably harvested.

At the same time, however, a debate has been growing about the extent to which fisheries science needs to shift its own focus, acknowledging that some of its practices — for example, the development of increasingly complex models of fish stock dynamics — and the way that the science and politics of fishing interact may need to be rethought.

A report published last year by Britain's House of Lords, for example, which dramatically highlighted the current crisis, said that scientists were partly to blame for the perilous state of fish stocks because their cautiously worded advice has been used by fisheries managers to continue the practice of overfishing (see *Nature* 379, 481; 1997).

In common with other areas where scientists have been asked to advise on potentially risky social practices, new models are now being explored of ways in which the various 'stakeholders', including both individual fishermen and representatives of the fishing industry, can become involved.

Scientists have been formally advising North Atlantic governments on fisheries management ever since the International Council for the Exploration of the Seas was set up in Copenhagen in 1902. Organizations such as ICES work by providing governments with assessments of fish stocks and by suggesting a 'total allowable catch' based

on the idea of a maximum sustainable yield, in other words estimating the quantity of fish that can be taken out of the sea without disturbing fish regeneration.

Yet the failure of scientists both to understand fisheries biology and to forecast several fisheries collapses has led to questions being asked as to whether the advice has been inaccurate, ignored or misunderstood — and whether it needs to change.

Advice irrelevant

One of those who believes that the nature of the advice has been largely irrelevant to effective fisheries management is John Beddington, director of the Centre for Environmental Technology at Imperial College in London, which houses an active fisheries research group as well as a consultancy that advises many of Britain's dependant territories on fisheries management.

Beddington says scientists should have recognized long ago that knowledge of the biology of fisheries is too underdeveloped to be used as a basis for advising governments on setting sustainable catch limits. Rather than wait until the state of knowledge becomes 'perfect', Beddington advocates

devising models on fisheries management that take account of — rather than try to eliminate — uncertainties in the knowledge of fish population biology.

Moreover, he says that setting catch limits has proved near impossible to regulate as it effectively needs officials to police each individual vessel. Beddington says there are more effective ways of managing fisheries, such as by controlling the size and fishing area of boats. Management on this basis, known as 'effort control', would eliminate the need to set catch limits based on uncertainties in the population biology of fish.

Some scientists, such as John Shepherd, director of the Southampton Oceanography Centre in the United Kingdom, believe that fisheries advice to governments has been too biology-based and needs to broaden out. He says scientists could also provide advice, for example, on the impact of different management policies on the fishing industry. But this, he says, requires drawing on the results of fisheries research programmes that encompass economics, sociology and anthropology, as well as biology, all of which are only now getting started.

Others believe that the pursuit of ever more accurate stock assessments has diminishing utility in countries outside the North Atlantic, where most of the world's fishing takes place. In these countries, such as China, Japan, India, Peru and Chile, there is proportionately more family and community-based fishing than commercial fishing.

Almost half of total world landings come from small-scale activity. Such fishing takes place along rivers, and fisheries on the whole are not managed on the basis of individual catch quotas, as they are in the European Union (see page 110).

Meryl Williams, director general of the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Fish wars: trawlers blockade Tarbert Harbour in Scotland in protest at restrictions imposed through the European Commission.

International Centre for Living Aquatic Resources Management in Manila, Philippines, believes that scientific advice on developing country fisheries needs to be clear and workable, as scientists have more influence among policymakers than fishermen, who tend to be poorer and less well organized than in the developed world.

But not everyone agrees that biology should lose its central position. David Cushing, former deputy director of the UK Ministry of Agriculture's fisheries laboratory at Lowestoft in Suffolk, who has been studying the dynamics of fish populations since the 1940s, still believes there is merit in 'the old ways'.

Cushing acknowledges that scientists

have made errors in their predictions, such as failing to forecast the Grand Banks cod collapse off Newfoundland. But he says that the lessons from such collapses, and the experience of decades-old research programmes, deserve a voice in the corridors of power.

Complex biology

Fisheries science — and thus science-based advice to governments on fishing policy — has until now concentrated on population biology, including trying to accurately assess the numbers of fish in the sea, as well as the rate and age they should be harvested to achieve sustainability. But the task has proved to be more complex than first imagined.

Scientists set sail in search of influence

Some fisheries scientists see a possible counterbalance to the fishing industry's influence among senior policymakers in the creation of an intergovernmental panel on the oceans, similar to the United Nations' advisory panel on climate change (IPCC).

Their belief is that fisheries science would be taken more seriously by policymakers if it were to receive the high-profile advocacy enjoyed by the intergovernmental climate panel, whose latest report has played a crucial role in forging agreement on the need for an international law to curb greenhouse gas emissions.

Lord Selborne, a member of the UK House of Lords select committee on science and technology, is one of those who believes that reports from an equivalent panel of fisheries scientists could eventually lead to an international law to prevent overfishing. "The climate convention would not have happened without the IPCC," he says.

Fisheries is the subject of two United Nations conventions. But there is currently no agreement to regulate fishing in international waters. Selborne also believes that such a panel, as with the IPCC, would command the attention of senior environment ministers, "instead of junior agriculture ministers" with less influence.

But not everyone agrees. Meryl Williams, director-general of the International Centre for Living Aquatic Resources Management in Manila in the Philippines, says that an international panel of scientists will not necessarily result in fisheries moving up the political agenda.

"Already, there is no dearth of science advice," she says, pointing out that climate change achieved its international profile partly through the work of environmentalist groups, which is only just beginning to happen in fisheries.

At present, two bodies exist to provide governments with advice on fisheries science. The International Council for the

Exploration of the Seas (ICES) was set up in 1902 to advise North Atlantic governments. Its Pacific equivalent (PICES) was set up in 1994. Canada is one country that has expressed enthusiasm for a new, high-level panel. But the British government has repeatedly rejected the need for such a body.

Joe Horwood, deputy director of the UK agriculture ministry's fisheries laboratory at Lowestoft, Suffolk, says the regional dimension to fisheries is stronger than with climate change. "North Sea cod is a regional issue," he says. "If the cod is wiped out, it won't necessarily affect cod in other parts." Horwood fears that a global panel "could be used by some to stall the development of regional agreements".

But John Shepherd, director of the Southampton Oceanography Centre, says fisheries has an important global dimension which, he believes, is difficult to establish through regional forums such as ICES. He also says that fisheries science has been "too inward looking" and needs more comprehensive peer review, which would happen through a large global scientific body along the lines of the IPCC. "Papers in journals tend to be rare. Research results remain within a closed community." Most research is published in official, unrefereed reports.

The UN Convention on the Law of the Seas, one of two conventions that govern fisheries, opened for signature in 1982 and was ratified in 1994. It established legal principles for navigation and conservation, and decreed that coastal states had jurisdiction over fish within 200 nautical miles (370 kilometres) of their shorelines.

The UN agreement on straddling stocks, signed in 1995, is an attempt to resolve disputes over rights to fish that migrate across the 200-mile zones. The Convention on Biological Diversity, which was opened for signature in 1992 and has since been ratified by 145 countries, also deals with fish conservation.

E.M.

The reproduction of fish, known as recruitment, is unpredictable. The natural year-on-year variation in recruitment in the North Sea can change by as much as "two orders of magnitude", says Cushing. He believes that understanding recruitment is the key to setting sustainable catch limits. (His 50-year efforts to try to address these questions are summarized in a recently published book, *Towards a Science of Recruitment in Fish Populations*).

Joe Horwood, deputy director of the UK Ministry of Agriculture's fisheries laboratory at Lowestoft, says that, despite some recent progress, recruitment remains a largely unsolved puzzle. A consensus has emerged that research needs to concentrate on the first few months in the life of a fish when growth and mortality rates are highest.

The numbers of spawning fish depend on changes to the type of food eaten by fish larvae; and how many larvae end up as food for other sea creatures. Their potential predation, in turn, depends on how fast they can develop; the faster they grow, the less likely they are to be preyed upon. Many die during the journey, known as larval drift, from their spawning grounds to their 'nursery' grounds.

The effect of climate change is yet another factor. Fish tend to spawn at the same time each year, but this may not always coincide with the availability of food. "If the spring algal bloom is too early, or too late, the fish miss out," says Horwood. He says there is "some evidence" to suggest that larvae grow where there is more food. But these are questions for which there are no clear answers yet.

In contrast to Cushing's argument on the importance of trying to unravel each strand in the fish life cycle and its interaction with other ocean constituents, Beddington advocates constructing models of fish behaviour which incorporate these uncertainties, instead of trying to eliminate them.

He points out that the successful 'Revised Management Programme' of the International Whaling Commission is based on such an approach, and research groups at Imperial College are now working on ways of adopting this method for fisheries management.

But science advice in fisheries policy-making faces a number of constraints. One is that it is inevitably at the request of governments. And while governments continue to ask for advice on how to set catch limits, says Keith Brander, a spokesman for ICES, organizations such as ICES will have to comply.

Brander says ICES, and its Pacific equivalent PICES, have long recognized there are alternatives to giving advice on catch limits, such as managing by effort control. He says that ICES has also embarked on projects encompassing fisheries economics. But, unless governments themselves manage their fisheries in a different way, they will continue to ask for advice on setting catch limits.

Ehsan Masood

Canada's cod leaves science in hot water

The need for scientists and fishermen to develop greater trust was dramatically illustrated five years ago by events surrounding the largest collapse of a cod fishery in modern times, that of Canada's Grand Banks off the coast of Newfoundland.

The fishing area has remained closed since then, with the loss of 40,000 jobs. Alan Sinclair, of the fisheries department's Gulf Fisheries Centre in Moncton, says recovery is taking longer than expected. "There's still not much sign of recovery in areas in the Gulf of St Lawrence and off Nova Scotia."

The blame for the collapse is usually pinned on flawed models — based in part on the misreporting of data — together with political reluctance to persuade fishermen to reduce their catches in line with scientific recommendations.

Before the collapse, scientists would set a 'total allowable catch' each year based in part on assessments of fish stocks from reported commercial catches, and in part on data from scientists' own research vessel.

Data supported opposing views

The two sets of data did not always tally. The scientists' stock assessment would invariably be lower than data from the fishermen. The explanation was simple: fishermen went to warmer waters where cod were found, while the scientists traversed a random path. As a result, however, fishermen would insist there were enough fish in the sea — a message that invariably had the ear of the policymakers.

Bill Doubleday, director general for science in Canada's federal oceans and fisheries department, says the department has addressed the problem of mistrust by involving fishermen in the evaluation of stock assessments, as well as in collecting data.

In what are called 'sentinel surveys', the fishermen follow a scientific protocol, catching a small amount of fish according to an experimental design, mainly with fixed gear. In addition, some fishermen using mobile gear carry out deeper surveys at their own expense. "I think it was mainly having fishermen participate in the reviews that has defused the issue. It has helped very substantially, because now they understand what went on," Doubleday says.

Tensions between scientists and fishermen have not been eliminated by this strategy. But instances of public disagreement have been significantly reduced. "We've changed our perspective from ten years ago, when there was a tendency to err on the side of the fishermen. Now if we are going to err we err on the side of maintaining enough fish in the water."

Yet there were other reasons for the collapse, and workshops have been held to examine the methods used in stock assessment at the time. Scientists say that here, too, many

lessons have been learned. "We have tried to correct them by dockside monitoring and audits and also to understand how our models are performing," says Sinclair. "There are also techniques used that don't use commercial data at all."

Another factor in the collapse is thought to have been the failure to understand the variable mortality rates of cod. The populations of Canadian cod were considered robust, as they would always bounce back after depletion occurred.

"The red flag wasn't waving as hard [as in Europe's North Sea] and for a brief period in the late 1980s and early 1990s there were attempts to phase in reductions in the fisheries over a period of years," says Sinclair. "By that time the evidence was really strong that there was a problem, and Canada closed the fishery down."

According to Doubleday, the Canadian cod stocks turned out to be less robust than had been thought. "Our level of exploitation was too high for those circumstances, even though it was nowhere near as high as in the North Sea."

Both Sinclair — the co-author of a recent paper (see *Nature* 385, 521–522; 1997) arguing that the rate of exploitation of North Sea cod is unsustainable — and Doubleday believe that the Grand Banks collapse cannot be compared to the North Sea.

Doubleday points out that the Canadian cod were in a very cold part of the ocean, with bottom temperatures of -1.4°C , while North Sea cod are in waters that do not fall below 4°C . Cod stocks are larger where water is relatively warm.

But they point out that catches are much larger in the North Sea than they were in Canada. North Sea cod tend to grow quickly,

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

The way it was: full catches, seen here in Ireland, are little more than memories in parts of Canada.

says Sinclair, but they are caught at age two, before they mature sexually. And if fish spawning should fail for two years in succession the fishery could be reduced rapidly.

"Some consider it amazing that the North Sea cod survive because the fish are taken before the age of four years and the fishing pressure is very high," says Doubleday.

"One thing they might do is increase the size at which they are caught," suggests Sinclair. "This could be accomplished by increasing the size of the net mesh. But a problem is that the catch includes whiting, which are smaller than cod, and these might be lost with a larger mesh." **David Spurgeon**

Overfishing is a global threat, warns FAO

The State of World Fisheries, the latest report from the United Nations Food and Agriculture Organization in Rome, could make even the most eager fish gourmet think again before placing their next order. More than two-thirds of the world's species — including favourites such as cod, lobster, prawn and shrimp — are variously being fished to capacity, overfished or are recovering from overfishing.

The organization predicts that world fish production will remain fairly constant at around 90 million tonnes per year up to 2010. But, with world population expected to reach 7 billion by this time, fish supplies will be hit by a 19-million-tonne shortfall, expected to be met by fish farming.

Eighty per cent of the world's marine catches are taken by 20 states. China, Japan and North and South Korea account for almost one-third of the world's total fish

consumption. Forty per cent of the world's 25,000-strong fishing fleet is in Asia, followed by 30 per cent in the former Soviet Union, 10 per cent in North America and 12 per cent in Europe.

China heads the international league for fish production. Fifteen million tonnes were produced in 1992, much of it for export. Japan is second at 8.5 million tonnes. But the Japanese eat more fish, reaching a total of 70 kilograms per person per year. Peru and Chile both catch more fish than the United States — 8.5 and 6 million tonnes respectively.

In the European Union, fish production increased from 8 million tonnes in 1984 to 9.5 million tonnes in 1992. Average consumption per person is now 22 kilograms — the same as in the United States and Canada, which together catch 7 million tonnes of fish per year.

Researchers bring fishermen on board

How is climate change likely to affect the abundance of marine animals? And can tensions between scientists and fishermen be reduced if the latter can become involved in research by collecting data during their work? Both questions are to be addressed by a new US research programme which aims to generate knowledge that will be essential for the establishment and maintenance of sustainable fisheries.

The GLOBEC (Global Oceans Ecosystem Dynamics) programme is funded by the US National Science Foundation, the National Oceanic and Atmospheric Administration and the National Marine Fisheries Service. It focuses on three sites: the Northeast Pacific, the Antarctic region, and the Georges Bank fishery in the Northwest Atlantic.

The most prominent is the Georges Bank study, as it will draw on data gathered both by research scientists and commercial fishermen. The idea is that such collaboration could forge ties between communities that have often been at odds, while eventually paving the way for more effective fisheries management strategies.

Search for fluctuations

Scientists are trying to see how populations of zooplankton, copepods (small marine crustaceans) and imperilled cod and haddock fluctuate in Georges Bank in response to varying physical and biological conditions. The effort will require extensive sampling of the waters of this and adjacent regions.

"We need a tremendous amount of information, gathered over a broad area and many years, which is where the fishermen come in," says Peter Wiebe, a biologist at the Woods Hole Oceanographic Institution in Massachusetts. "They're out in places we don't have the resources to get to, and it makes sense to enlist their services." Participating fishermen might either collect data during a normal fishing run or go out on special research trips.

The idea has been welcomed by New England fishermen, who have been hit hard by the collapse of local fish stocks as a result of overfishing. "Fishermen are quite receptive to this opportunity," says Rollie Barnaby, a former fisherman who now works for the Sea Grant College Program at the University of New Hampshire (UNH). "Not only can they learn something and contribute to science, they can also earn a little extra money."

Craig Pendleton of Saco, Maine, is one fisherman who is eager to get involved in the



Woods Hole's research vessel *Oceanus*.

project. Owing to government restrictions, Pendleton will only have 88 days this year to fish for cod, haddock and other so-called 'groundfish', leaving him plenty of time for research-related endeavours. He believes that fishing vessels and their crews could offer an economical means of collecting data at sea. "We have a lot of experience in the water and are used to operating as cheaply as possible," he says.

But the goals of the Georges Bank project extend beyond the financial concerns of fishermen and scientific researchers. An equally important objective is to improve relations between the two groups and to enhance understanding on both sides.

"Communications between scientists, government regulators, fishermen and fish processors have never been good," Wiebe says. "Fishermen either don't understand what we're trying to do or they don't believe what we tell them." He hopes this adversarial relationship will turn into a "collaborative process," with scientists learning from fishermen, and fishermen, in turn, becoming better schooled in scientific methods.

Agreement is the starting point

Ann Bucklin, a biologist at UNH who heads the New Hampshire-Maine Sea Grant College Program, says: "Fishermen need help understanding the biological reasons for the regulations they consider themselves to be 'inflicted' with. The way to fix that problem is to get fishermen involved in the collection of the data that provide the basis for fishery management strategies and regulations."

Scientists frequently rely on fishermen to install or recover monitoring equipment, but the Georges Bank project represents "the first concerted effort in this country to involve fishermen in a major oceanographic research program," according to Bucklin.

Although 70 scientists are participating in the project, the budget to involve fishermen is small—limited, so far, to a single year of funding (1997) and a total of just \$30,000. The money will barely cover the three research forays planned for fishermen this spring, summer and autumn.

But a proposal will be submitted to expand the budget for 1998. "There may be creative ways of funding this activity, if it turns out to be a good way to go," says Wiebe. Apart from direct grants from the federal government, other possibilities might include tax credits, contributions from the fish-processing industry, and extra fishing days granted to those taking part in the research.

Wiebe sees large potential for cooperative ventures of this nature. "Three years from now, GLOBEC Georges Bank will be completed, but the need for this kind of information will still be here," he says. **Steve Nadis**

Too hot for a homecoming?

How far should scientists go in deliberately attempting to reverse the disappearance of stocks of fish of potential commercial value? The issue has been raised by a project intended in part to establish whether climate warming was responsible for eliminating from English waters Europe's only freshwater member of the cod family.

The fish in question is the burbot (*Lota lota*), which is still abundant in other, less temperate, regions of Europe. In a pilot project involving 100 specimens imported from the Czech Republic, Jim Reader, a lecturer in life science at the University of Nottingham, is hoping to discover whether it might be reintroduced.

Some fear that climatic conditions could be inappropriate for the fish, which disappeared from Britain 25 years ago. Alwyne Wheeler, former keeper of fishes at the Natural History Museum in London, says that such a reintroduction would be "morally improper". Wheeler claims that the British burbot was doomed to extinction about 12,000 years ago, when the final

retreat of the ice cap and rising sea levels left it stranded in Britain.

Several freshwater fish, such as the arctic char and races of brown trout, have disappeared from many parts of England since the Industrial Revolution of the eighteenth century. In most cases, this has been blamed on local causes, such as sewage or industrial activity.

In contrast, the burbot, which some say was abundant in lowland rivers in eastern England (and of considerable commercial importance), has disappeared completely, following a few final sightings around 1970.

Unlike most other British freshwater fish, the burbot spawns in mid-winter at temperatures slightly above freezing. Some scientists, such as Wheeler, believe that English mid-winter temperatures are now too warm for the burbot, a result of natural and man-made climate changes that are widely believed to have caused a rise in average global temperatures. Reader accepts that this is a reasonable hypothesis—but one that has yet to be proved either way.

Aquaculture: a solution, or source of new problems?

Until recently, aquaculture — the selective breeding and raising of fish in 'fish farms' — promised to resolve problems caused by the world's growing shortfall in marine fish supplies. But increasing awareness of its unwanted side-effects means that aquaculture itself is also heading for troubled waters, particularly in the developing countries.

Intensive fish farming in India and China, for example, the world's largest aquaculture producers, has led to pollution and the spread of disease. Evidence is also emerging that the promotion of aquaculture — 84 per cent takes place in Asia — is causing significant unemployment among those engaged in traditional fishing.

Last month, India's Supreme Court dismissed a call for a judicial review of a decision last December demanding the closure of virtually all commercial shrimp farms in five coastal states by 15 April, and the payment of compensation to employees made redundant.

The decision includes a ban on all aquaculture in mangrove swamps, estuaries, wetlands, and on public land, as well as a prohibition on converting agricultural land into shrimp farms. The court had ordered the closure after a long campaign by environmentalist groups protesting in particular at the contamination of land and drinking water from the discharge of toxic effluent.

Responsible for increased yield

Despite growing problems, the United Nations' Food and Agriculture Organization (FAO) remains committed to its prediction that a 19 million-tonne shortfall in global fish supplies by 2010 will be filled by aquaculture. Aquaculture currently contributes

just under one-fifth of world fish consumption, and is generally credited as being responsible for most of a 3 million-tonne increase in world fish consumption between 1990 and 1993.

Following FAO advice, and with the help of funding from the World Bank, most of Asia's leading fishing states have recently been taking major steps to develop aquaculture, which absorbs virtually all funding for fisheries research in Asian countries.

Aquaculture research in the Philippines, for example, received more than 80 per cent of fisheries research funding during the 1980s, according to Resource Allocation to Fisheries Research in Asia, a report published by the Asian Fisheries Society in Manila.

Aquaculture concentrates on three classes of fish: finfish, crustaceans and molluscs. Sixty per cent of production is in inland areas, with the rest along coastlines. And, rather than develop small-scale aquaculture for domestic consumption, countries have opted to invest heavily in intensive aquaculture for species such as shrimp, as a ready source of foreign exchange from exports.

Intensive aquaculture involves stocking vast ponds with large quantities of fish, food and chemicals. "To raise one tonne of shrimps requires 50 to 60 million litres of water," says Biksham Gujja, head of freshwater policy at the World Wide Fund for Nature in Geneva.

But this inevitably affects water supplies. Shrimp farms, for example, tend to be located near the coast; as a result, around 2.5 million cubic metres of effluent are discharged daily by farms on the east coast of India, adds Gujja. "This works out at about 15,000 litres

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Paradise lost: prime paddy land in Thailand has been destroyed by pollution from shrimp farming.

of effluent per kilogram of shrimp."

Evidence presented by environmentalist groups at the Indian Supreme Court hearing showed that pollution from shrimp farms had affected drinking water, cropland, mangroves, groundwater aquifers and coastal fisheries. But whether such 'modest quantities' can meet the shortage in fish supplies remains to be seen.

One problem is the pressure that comes from a desire to seek quick profits. Malcolm Beveridge, head of the environment unit at the Institute of Aquaculture at the University of Stirling in Scotland, says many intensive shrimp farmers operate a 'flash and burn' policy by which farms are built, operated and abandoned within a short time.

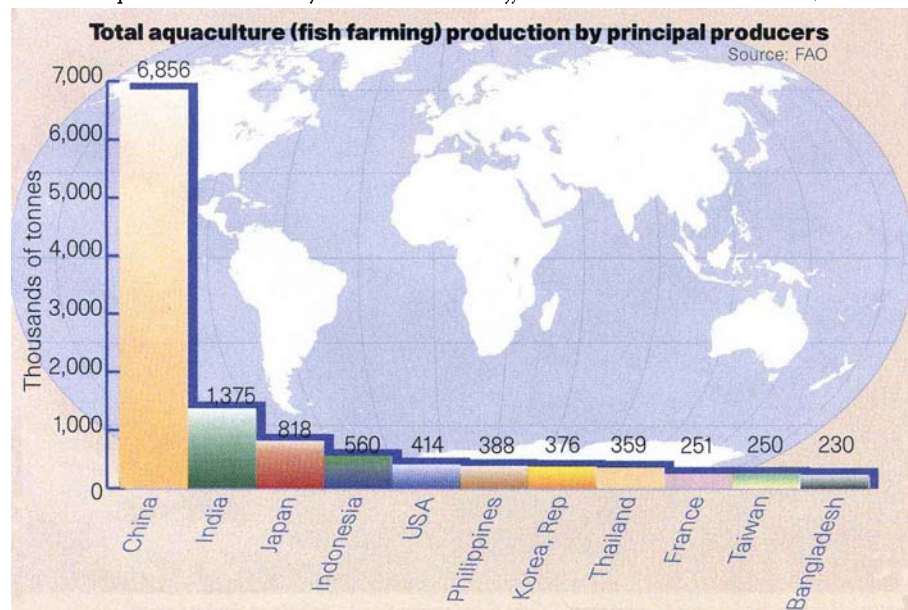
Operations being closed down

Beveridge says that farm owners circumvent the disease problem by closing down their operations after four or five years and moving to a new location. "Thailand passed a law against this practice. But it doesn't seem to be working," he says.

Another concern is the impact on marine fishing from intensive aquaculture, particularly in countries where fishing by local communities is a direct source of food for families. "Aquaculture has a lot to offer," says Meryl Williams, director-general of the International Centre for Living Aquatic Resources Management in Manila. "But there needs to be a careful look at the winners and losers."

Beveridge believes that aquaculture can still make a significant contribution to fisheries production, providing it is done sustainably. Sustainable aquaculture, he says, is ideally performed inland, not near the coast. Also, it should use "only modest quantities of fertilizer [for the production of fish food], and water".

E. M.



Fishing by numbers reveals its limits

Fifty years ago, a single cod was large enough to feed a family of four or five. Today, it is barely enough for one, says Lord Perry of Walton, a member of the UK House of Lords select committee on science and technology, who grew up on the east coast of Scotland where much of the UK fisheries industry is based.

The root of the problem is overfishing — including the practice of taking fish before they are old enough to breed and lay eggs — which is considered to be the most critical issue in global fisheries today. But attempts to persuade fishermen to reduce their catches have spawned the recognition that a softer approach may be needed to persuade them to listen to scientists, who they often distrust.

Overfishing is believed to have been an important factor in the Grand Banks collapse off the coast of Newfoundland (see page 107), and the American Georges Bank collapse, and is now threatening cod stocks in Britain's North Sea (see *Nature* 385, 521; 1997). It is expected to continue, as no country fishing the Atlantic or Pacific appears to have found a solution that satisfies the fishing industry.

Fishing in European Union waters is regulated by assigning catch quotas for individual member states under the Common Fisheries Policy. In force since 1972, the policy has proved near impossible to enforce. Fishermen often exceed their quotas, and under-reporting is rife. The latter has the additional effect of introducing errors into the many fish databases to which fishermen are required to report their data.

Realpolitique makes its mark

Keith Brander, a spokesman for the North Atlantic fisheries scientists' advisory council, ICES, in Copenhagen, says the roots of the Common Fisheries Policy lie in an attempt by the then European Economic Community to seek recognition from the Soviet Union, which also had significant fishing interests. One way of achieving this was to meet the Soviet Union on equal fishing terms. "It was more a case of *realpolitique* than a desire for conservation," says Brander.

In its report *Fish Stock Conservation and Management*, published last year, the Lords committee called for an immediate 30 per cent reduction in 'fishing effort' — a measurement based on the size of a fishing boat and the area in which it can fish — throughout the European Union.

The Lords recognized that a reduction in fishing effort must be Europe-wide; if a single country decided to reduce the size of its fleet, neighbouring boats would happily fill in the breach if the regulations did not apply to all. A limited boat decommissioning policy has been in operation. But, as the Lords wryly noted, the industry tends to shut down

its most inefficient trawlers, making little difference to the overall catch.

To overcome the fishing industry's distrust of researchers, the Lords report called for panels of scientists, fishermen and fisheries managers to regularly sit down and discuss issues. It also urged scientists to be less ambiguous in their advice, which is often used by politicians to delay measures to reduce fishing.

Lord Selborne, a member of the Lords committee and a former chairman of the Natural Environment Research Council, says that scientists need to broaden the focus of their perspective to take account of the social and economic implications of reduced fishing, if their advice is to have an impact.

This view is supported by Michael Healey who leads a research project at the University of Rhode Island that is comparing the use of science in fisheries management in the United States and Canada. He says any successful method must give fishermen "a sense of ownership", and that fisheries authorities should not over-regulate, "as this simply encourages fishermen to disobey the rules".

Healey says the Georges Bank fishery off New England is a classic example of over-regulation. Initially there was a single, annual quota, but this was reached within three months, mainly by the large vessels. Owners of smaller boats complained that they had been deprived of access because of bad weather.

After discussions with biologists, the quota was increased. But, says Healey, a combination of large and small boats cleaned up within three weeks fish that were meant to have lasted three months. "The managers then imposed weekly limits, and then individual vessel quotas. It was really getting bizarre. But nothing seemed to work," says Healey.

A similar situation occurred in Canadian fisheries management in the 1970s. "Initially,

managers assigned separate quotas to inshore and offshore vessels. But in a few years the inshore boats fished both sets of quotas." Managers responded by allocating quotas according to vessel classes, but found that larger vessels were impossible to control.

Failure of restrictions

The next stage was to restrict fishing to certain areas. This too failed, as did an attempt to impose restrictions according to species. "If you try and close off areas to fishing, people just do it anyway, as there's no way of checking unless you have a policeman on every vessel."

Healey adds that species restrictions invariably led to a high quantity of discards, as nets cannot differentiate between species of similar size.

When fishermen are asked why they disregard the regulations, they say that the rules are "too damn complicated, and they didn't understand them", says Healey. He insists that regulations "must be kept simple", even if this means that some classes of fishermen are affected more than others. "If you try and give every fisherman what he wants, it gets too complicated." He adds that regulations must find a way of delegating at least part of the responsibility of fisheries management to the fishing industry.

In south west Nova Scotia, for example, fishermen have agreed to design and administer their own catch quotas, and have been given the right to penalize rule-breakers. Some fish producers, meanwhile, are beginning to take the initiative themselves. Unilever has launched a consumer scheme with the World Wide Fund for Nature in which it will only buy sustainably harvested fish, the proof of which will appear as an 'eco' label on its products.

E.M.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Factory fishing: Soviet trawlers, such as this in the Barents Sea, have helped deplete stocks.

Plant modification needs more discussion

Sir—Donald Ort (*Nature* 385, 290; 1997) argues that “[t]he interspecific genetic modification of foods is not inherently new” and that “the more the focus is kept on safety of the product for humans and the environment, with decisions being made on the basis of the soundest scientific findings rather than on novelty, the better the result should be for consumers and the environment”.

The idea that genetically modified organisms (GMOs) are the same as conventionally bred plants, and therefore impose no extra risks, is at odds with the intellectual property protection — patenting — that is sought for GMOs. Claiming European patent protection for GMOs actually depends on proving to the European Patent Office in Munich their novelty, non-obviousness and usefulness. Genetic engineers cannot have it both ways depending on whether they are trying to allay the public's fears of GMOs or to obtain a patent.

The issue of patenting GMOs and the products and processes of genetic engineering will be discussed in the European Parliament in the next few months as the European Draft Directive on the Protection of Biotechnological Inventions (note the word ‘Inventions’, not ‘Discoveries’). So far, there has been relatively little discussion of the many aspects involved in patenting plants, parts of humans and other animals, not least the ethical aspects.

Recent questions about the importation of genetically modified soybean into Denmark and the rest of the European Union has concentrated on whether such produce should be marked. This is essentially a discussion about the freedom of individuals to decide the food they wish to consume, irrespective of the safety issue, and is thus an issue of individual ethics. The purpose of the draft European Directive is to allow the patenting of such things as “elements isolated from the human body” (Article 3). This would almost certainly include an individual's DNA if that DNA were used to produce a drug by a company — and the company would hold the patent. Further, there is nothing in the directive to prevent even the patenting of an *in vitro* fertilized human embryo — only “the human body and its elements in their natural state” are excluded from patenting (Article 3). Other patentable objects would include “biological material including plants and animals” (Article 4). Exceptions to patentability seem to be made for “methods of human treatment involving germ line therapy” (that is, genetic modifications of sperm or eggs) and where “processes for modifying the genetic identity of animals cause them unnecessary suffering” (Article 4). However, exploitation of an invention cannot be prevented in the directive solely on the grounds that such an exploitation is illegal (Article 9).

When the European Parliament rejected

the Draft Directive on the Protection of Biotechnological Inventions in its first round in February 1995, members of the European Parliament were widely applauded for their responsibility and sound judgement. Their decision was based on the insight that the patenting of life has implications that reach far beyond the technological and economic domain. It touches wide and deep dimensions of ethics, ecology, culture and religion. This is not to deny that inventors need rewards, but it must be remembered that a patent is a monopoly right to control the exploitation of an invention and is given to a private organization or person by society. This monopoly right must therefore respect the wider ethics and values of that society. A European Commission directive is not the way to come to a just balance between these interests. The new directive should be similarly rejected pending a full discussion of the rights, responsibilities and means of intellectual property protection within our society. The decisive arguments against the first version of the directive, raised in the parliamentary debate, are still valid and urgently need to be raised again with the present version, which still does not properly address issues of concern.

John R. Porter

Department of Agricultural Sciences,
Royal Veterinary and
Agricultural University, Agrovej 10,
2630 Taastrup, Denmark
e-mail: john.r.porter@agsci.kvl.dk

Plutonium disposal

Sir—As a recent leading article points out (*Nature* 384, 599; 1996), much discussion is going at present (mostly in the United States) on the policy for the disposal of excess plutonium from nuclear weapons.

An international conference, “Utilization/disposal of excess fissile weapon materials”, organized by the Landau Network under the auspices of the Italian Ministry of Foreign Affairs and Unesco in Como, Italy, in March 1996, considered the problem of the value of weapons-recovered plutonium as a reactor fuel.

Many of the invited scientists agreed that the question is essentially irrelevant for two reasons. The first is that utilities in several countries — many of them operating on strictly market rules — choose today to reprocess their fuel, partly to facilitate waste disposal and also to use the recovered plutonium in MOX fuel for their light-water reactors (LWRs), rather than disposing of it

in some other way. The second is that the permanent disposition of weapon-derived plutonium as a fuel has a value that should not be measured in strictly monetary terms. The conclusions of the Como Conference are of course unofficial, but at the same time they present an independent scientific look at the problem.

On the other hand, as a follow-up of the Moscow Nuclear Safety and Security Summit in April 1996, an international meeting on “Safe and effective management of waste fissile material designated as no longer required for defence purposes” was held in Paris in October 1996. The meeting was technical and focused mainly on the management of the excess weapons plutonium. The experts concluded that there are certain options that offer prospects of early progress towards the non-proliferation and other objectives set by the Moscow Summit. These are approaches involving consumption of weapons plutonium as MOX fuel in LWRs. Immobilization is also a viable,

complementary option that may be suitable for some countries in their national plutonium disposition strategy.

The experts at the Paris meeting concluded, therefore, that the MOX option, under appropriate non-proliferation conditions and international verification, offers greater promise for the management of the excess weapons plutonium. The immobilization can be an appropriate temporary component of the plutonium management strategy of some concerned nations.

We also believe that these reflections on different scientific and technological problems of nuclear security should be continued and supported by international and intellectual organizations, including Unesco, which is developing its Culture of Peace Programme with a strong scientific component.

Maurizio Martellini

Department of Physics,
University of Milano,
Italy

Japanese grant system needs more resources

Sir— Shinji Yokoyama's letter reveals the perception of a researcher associated with the Japanese grant system (*Nature* 384, 402–403; 1996). But your readers should be told of recent reforms and the need for more resources. (I write as chairman of the committee on grants-in-aid for scientific research in the Science Council of the Ministry of Education, Science and Culture.)

Despite the recent rapid budgetary increase, there is still a gap between the needs of researchers and the available resources for grants, which for the fiscal year 1997 (which starts on 1 April) amount to roughly US\$1 billion. This may all be spent on direct research expenses other than salaries and overheads, but when divided among 35,000 annual awards it averages no more than US\$25,000, with an average support period of 1.5 years.

Starting from fiscal year 1997, however, the size and duration of support for small and medium-size grants will be enlarged in an attempt to provide a more stable research environment. To avoid too negative an effect on the success rate, a substantial budgetary increase is called for

in the next few years.

As for the proposal review, the 2,000 or so researchers involved are doing their best to provide fair and objective reviews. For the vast majority of small and medium-size grants, there is a two-round review system. For large-scale grants, outside referees are invited to comment, and applicants are interviewed. Although Yokoyama says that "many of the applications are outside the speciality of the reviewers", the reviewers are in fact chosen from a list of researchers recommended by academic associations in each research area. They are appointed for a single term of two years. After the term is over, the names of all reviewers are made public.

Since fiscal year 1996, a summary of comments from reviewers has been sent to failed applicants for large grants. From fiscal year 1998, the number of applications that first-round reviewers are asked to mark will be reduced. A renewal system of grants will be adopted soon, but cautiously, as this would also negatively affect the success rate of new proposals without a budgetary increase to finance it.

More reforms will follow because a major review exercise is in progress on the evaluation of the government's research-related activities.

A further increase of the grants would not result in a greater concentration of

resources in 'rich' universities. Although the 'priority research area' was criticized for such concentration, large numbers of researchers participate in one 'priority area'; naturally, a great majority of them come from other universities. The planned increase in the grants would fund the reforms and is expected to contribute to the creation of a more desirable research environment.

Fumimaro Takaku

*Jichi Medical School,
3311-1, Yakushiji,
Minami Kawachi-machi,
Kawachi-gun,
Tochigi 329-04, Japan
e-mail:ftakaku@jichi.ac.jp*

Alphabets and citations

Sir— There is an alternative explanation for the variation in citation depending on an author's alphabet position that Tom Tregenza refers to (*Nature* 385, 480; 1997). Variation may not be related to position in reference lists but rather to a real variation in productivity or originality of publication.

In our field (sedimentology) it is noticeable that this effect is exaggerated by several unrelated Allens publishing prolifically, and one in particular, J.R.L.

New Swedish solutions for purifying peptides of any source using any technique



INTRODUCING:

- ten new reversed phase chromatography columns
- two new ion exchange chromatography columns
- a new size exclusion chromatography column
- one completely new system for peptide, oligonucleotides and other biomolecules (AKTA is the Swedish word for real, it's pronounced eckta).

Allen, being very highly cited. One explanation for high citation rates for people near the start of the alphabet may be childhood conditioning. Children with names near the start of the alphabet are frequently called on by teachers to do things first and consequently not only get more attention but also more criticism when they fail (as they will be the first in the class to be seen to fail). Consequently, these people may either get used to failing in public or expect to do things first and well. Some of these children will therefore grow up with a heightened desire to achieve.

Less recent correspondence in *Nature* proposed that success rate in examinations, and in sport, depended on birth date being near the start of the school or sporting year. One might therefore predict that the children most likely to do well are those born early in the school year to families with surnames near the beginning of the alphabet.

In conclusion, it would be unfair to penalize over-achievers at the start of the alphabet by the modifying factor suggested by Tregenza. But then we would say that, wouldn't we?

Jan Alexander

Julian Andrews

School of Environmental Sciences,
University of East Anglia,
Norwich NR4 7TJ, UK

No knife needed

Sir — I was disappointed to read that you feel that, as South Africa seeks to reform its science base, it would be a mistake if the research of the Atomic Energy Corporation (AEC) were to escape the knife (*Nature* 384, 1; 1996).

In 1990 the corporation put in place, of its own volition, a ten-year plan to reduce its dependence to the level sufficient only to fund the national requirements of the state.

In the sixth year of the plan, state dependence has been reduced by 72%. This has been achieved by closing down research programmes not deemed relevant to the new South Africa and by commercializing many of the unique technologies developed during the period of political isolation.

Not that these cutbacks have been without loss: it is estimated that South Africa lost 22,000 man-years of high-level scientific experience in the process. On the other hand, the commercialized technologies are now being manufactured (without state subsidy) and marketed both locally and overseas. Sales in the present financial year (excluding traditional nuclear fuels) will be R148 million, of which 35% will be to international markets.

During this transition, the corporation has made some major achievements. In

1989–90, for example, the AEC dismantled the South African nuclear deterrent, and this was subsequently verified and favourably commented on by the International Atomic Energy Agency. More than a million patients benefit annually from radio pharmaceuticals, mainly molybdenum-99 for nuclear medicine, prepared in the corporation's laboratories. The markets are Australia, China, India, the European Union, Israel and sub-Saharan Africa. The corporation is leading research towards the single step enrichment of uranium by molecular laser isotope separation techniques. In addition, both the US and Russian space agencies are evaluating precooked foods irradiated by the AEC for long shelf-life.

The AEC has emerged from this continuing transformation with a reputation in South Africa as an applied research organization with a proven competence to deliver technologies to the market. Our recent annual report shows that the knife is now neither necessary nor deserved. The AEC has for many years now had an open door policy, and we welcome visitors. Our website address is: <http://www.aec.co.za>.

L. D. Hyslop

Atomic Energy Corporation of South Africa
Limited, PO Box 582,
Pretoria 0001, South Africa

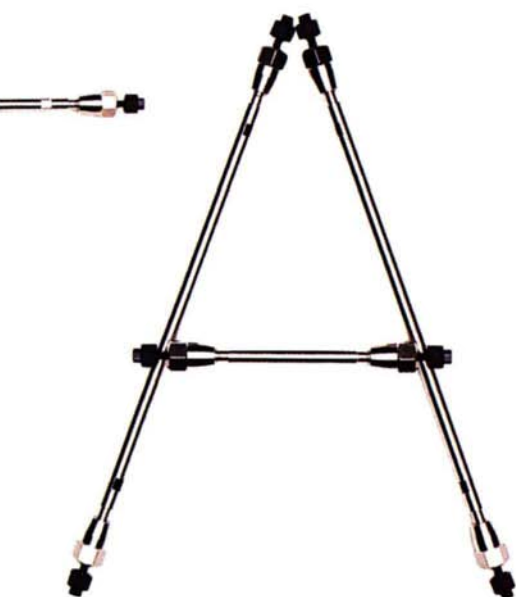
Are you working with natural peptides? Synthetic peptides? Recombinant peptides? Peptide fragments? Or all of them? Whatever peptide you work with, your options for purifying them just increased.

With ten new columns for reversed phase chromatography, you're nearly certain of finding the selectivity you need in our extensive range. All of these new RPC columns deliver high resolution; combined, they'll take you from purification and analysis to peptide mapping.

We can support you with advice and solutions for other peptide purification techniques as well. Are you separating peptides with poor solubility? Our new size exclusion column withstands high pH and solvents. Do you need an extra technique to help you with a difficult-to-separate peptide? We have two new ion exchange columns that permit very high resolution and withstand high pH ranges.

What's more, all 13 columns are supported by ÄKTA™ purifier—a revolutionary new purification system for peptides, oligonucleotides and other biomolecules. Its preset protocols let you resolve all major purification tasks quickly and easily. Its control system lets you instantly transfer your methods to purification systems at all scales.

Want to know more about our peptide purification solutions? Call us: 1 (800) 526 3593 from the USA; +81 (0)3 3492 6949 from Japan; or +46 (0)18 16 50 11 from Europe and the rest of the world; or meet us on the Internet at <http://www.biotech.pharmacia.se>.



**Pharmacia
Biotech**
Uppsala, Sweden. (And the rest of the world)

Public support for medical research

Sir — A majority of the American public — not just a select few in the Congress (*Nature* 385, 375; 1997) — support doubling funding for medical research in the United States. Public opinion polls, commissioned by Research!America and conducted by the Charlton Research Company between 24 January and 6 February found that 60 per cent or more of the respondents in Alaska, Louisiana, Ohio and Wisconsin favour the idea of doubling total US national spending on medical research by the year 2002.

Not by coincidence, similar polls conducted in the home states of Senator Connie Mack (Republican, Florida) and Senator Phil Gramm (Republican, Texas) showed the majority of their constituents stating that medical research should be made a higher national priority.

The suggestion by Senator Arlen Specter (Republican, Pennsylvania) that his Senate colleagues should “look toward alternative methods of financing” as well as giving the National Institutes of Health a yearly boost is also supported in previous Research!America polls (both national and statewide). More than half of the public agreed to pay \$1 more in taxes per week, \$1 more for each prescription drug, and/or \$1 more per week for health-care insurance if they could be assured the funds would be used to support additional medical research.

Mary Woolley

Research!America,
1522 King Street,
Alexandria, VA 22314, USA
e-mail: Researcham@aol.com

Passing the buck

Sir — We read with concern of yet another scandal in science — the case of fraudulent data in five papers on which Francis Collins, head of the Human Genome Project, is the senior author (*Nature* 384, 6; 1996).

Although we sympathize with Collins, we find it strange that a scientist initially willing to assume the senior authorship of a paper is later able to lay all the blame on his coauthor when things go awry. Almost inevitably, it is senior authors who get most of the credit for important papers (and sometimes the accompanying Nobel prizes), so why not the blame also? Even more mysterious is Kenneth Ryan's assertion that “it could happen to anybody”. Surely a scientist closely involved in the work published under his or her name should be able to prevent serious deception — after all, in the Collins case, the fraud was uncovered by an anonymous referee

without intimate knowledge of the work.

Reflecting on his own case, David Baltimore has suggested that scientists are “much too quick” to support allegations of scientific misconduct. Maybe the reason is that many scientists know all too well how large laboratory research occasionally operates — with senior faculty members involved only peripherally in continuing work — and extrapolate this general knowledge to controversial cases.

Standard criteria of authorship do not always translate into practice. How else can one explain faculty members who produce numerous papers each year while simultaneously travelling to meetings, delivering lectures and perhaps also consulting, running companies or managing large research organizations? How can one possibly stay abreast of research in the laboratory in such circumstances? It is often difficult for graduate students or postdocs to remind the faculty involved that their grant money does not automatically entitle them to coauthorship; and the name of a senior scientist on the title page may in any case assist a hard-working young researcher in getting a manuscript published.

We believe that authorship must imply a substantial intellectual contribution to a paper or significant participation in the actual research. Indeed, we would like to suggest the convention that a faculty member in charge of a laboratory be acknowledged separately in papers (for example as “Authors X and Y, under the guidance of Z”) unless he or she has made direct and clear contributions to the research being reported.

Ambuj Sagar

Paul de Sa

Science, Technology and Public Policy Program,
Center for Science and International Affairs,
79 John F. Kennedy Street,
Cambridge, Massachusetts 02138, USA
e-mail: ambuj_sagar@harvard.edu
paul_de_sa@harvard.edu

Novel paper titles

Sir — Many years ago, my colleague Dr Stuart Warren expressed the opinion that the word ‘novel’ in the title of a paper could usually be replaced by ‘yet another’.

Apply the Warren criterion to the data published in *Nature* by Friedman and Karlsson (385, 480; 1997) and all becomes clear: as the rate of publication increases, there are bound to be many more papers of the ‘yet another’ kind. True novelty remains as elusive as ever.

Jeremy Sanders

University Chemical Laboratory,
Lensfield Road, Cambridge CB2 1EW, UK
e-mail: jkms@cam.ac.uk

The earliest Americans

Sir — Several of the assertions made by Paul Bahn in his review of *American Beginnings* (*Nature* 385, 128–129; 1997) are debatable, and as editor of that book I should like to comment on them.

I do not know whether there is a consensus about when the first people entered America. There are assuredly those who see it as much predating the earliest certain evidence for the Clovis people, the first generally accepted American population. Here, as with other such problems, one simply follows the evidence and the thought that seem most persuasive. My characterization of Clovis as the “founding population” can hardly be seen as “astounding”. It is perhaps the most widely held view.

The “numerous” earlier radiocarbon dates from South America are simply *not* “widely accepted”, but then neither are they so numerous. The three sites discussed were chosen as examples because, in recent years, they have been held by their advocates to be among the most convincing cases of pre-Clovis occupations. They have been the subjects of much discussion. But to term my “dismissal” of these sites “ill-informed” in the case of Pedra Furada, “unworthy” in that of the Meadowcroft dates and presumably at best inattentive in the instance of Monte Verde — rejecting its “widely accepted date of 13,000 BP” and ignoring “the indisputable human artefacts and features associated with its possible date of 33,000 years BP” — would seem to call into question the credibility of the reviewer rather than that of the reviewee.

That dismissive attitude is said to “permeate even [my] treatment of the Beringian material”. The single cited example is “...the probably very early Russian site of Diring, [which the editor set aside,] claiming it does not ‘have any bearing on the subject of this volume’”. Let it be noted that the age of Diring is variously estimated at 2 million to 500,000 years or less. The propriety of excluding it from a discussion of American origins seems obvious.

(The Guthrie article was included because it presents a unique and invaluable depiction of late Pleistocene large mammal assemblages in interior Alaska. The article is old; the data remain current.)

Finally, I submit that the “balanced and informed view” is in fact here. That Paul Bahn thinks otherwise may reflect a rather superficial familiarity with the problem and, perhaps, accords with his finding the array of archaeological evidence presented, “a dull read”.

Frederick Hadleigh West

Peabody Essex Museum,
Salem, Massachusetts 01970, USA

Woodstock of physics revisited

Ten years have passed since the now famous American Physical Society meeting that heard the first breathless accounts of high-temperature superconductivity. Now, in calmer times, practical applications are emerging.

Paul M. Grant

Snap quiz: who can tell me the winner of the 1987 Super Bowl? Not most physicists, I suspect, for whom it was certainly eclipsed by two events of far greater consequence that shared the early months of that year. One, the discovery of Supernova 1987A, perhaps portended the other: the announcement of superconductivity above liquid-nitrogen temperature on planet Earth — a dream fulfilled for many condensed-matter physicists like myself, whose careers had orbited around this elusive star.

The successful sighting¹ fell to W. K. Wu and C. W. (Paul) Chu and their teams of students and postdocs at the Universities of Alabama and Houston, following only five months after the publication in autumn 1986 by Georg Bednorz and Alex Müller² at IBM Zürich of their discovery of superconductivity in a previously unexplored class of compounds, the layered copper-oxide perovskites.

The 'inside' story of the hectic interval between the first week in January 1987 — when an announcement of the confirmation of Bednorz and Müller's discovery first brought 'high-temperature superconductivity' to wide public attention — and the week of the American Physical Society's March meeting, remains to be told. Suffice it to say that this period, and the last three months of 1986, were replete with incredulity, credulity, excitement, secrecy and a sense of immediacy in competition with one's peers, all of which resulted in, frankly, a substantial amount of intrigue and suspicion. All who participated surely came to understand, if they had not done so before, that physics is not only a science but, perhaps more significantly, an



Rising stars: Müller and Chu with Shoji Tanaka (right), whose Tokyo laboratory provided one of the first confirmations of Bednorz and Müller's discovery.

intensely human pursuit — something they do not teach you in graduate school.

The programme of the March meeting, held each year in a different US city, is 'cast in concrete' early the preceding December; thereafter, an absolute policy of no alterations prevails. By the deadline of 5 December 1986, for the 1987 meeting at the Hilton hotel in New York City, only one abstract had been accepted on the new materials: "Specific heat of Ba-La-Cu-O superconductors" by Rick Greene and his collaborators at IBM Yorktown. But the explosion of results that appeared in the new year prompted the meeting's organizers to take an unprecedented step. Brian Maple of the University of Cal-

ifornia, San Diego, was asked to put together a special post-deadline evening session devoted entirely to the discovery.

All those wishing to report results would be granted five minutes each, in order of the arrival of their request to take part — and did the requests rain in, reaching a downpour in the two weeks before the meeting, as confirmations of the Wu-Chu measurements were made. All in all, 51 presentations were to be given throughout the evening and early morning of Wednesday and Thursday, 18 and 19 March. That memorable and riotous session was to become our "Woodstock of physics", so named in honour of the village only 50 miles north where, in an obscure farmer's muddy field in 1969, the rock concert occurred that defined a generation of youth the world over.

Opening act

A few personal observations and anecdotes may help to convey the colour of that week in midtown Manhattan. Excitement was running high even before Wednesday night. On Monday, the opening day, the press were already beginning to catch some of us to be interviewed. That noon my colleague Ed Engler and I went to lunch at a nearby Brew 'n' Burger and found Alex Müller sitting by himself in a corner booth, attempting to escape the turmoil at the Hilton. At the time he was not yet widely recognizable to those attending the meeting or to the press — a situation that would soon change.



Fever pitch: the room filled to overflowing with physicists eager for news of superconductivity.



Main attraction: Müller's appearance at the meeting was greeted with enthusiasm by the crowd.

On Tuesday morning, the 17th, as the Saint Patrick's Day parade began only a block away, Greene gave, to an overflow audience, the only talk on high-temperature superconductivity printed in the meeting bulletin. He began by introducing Müller to the crowd, who responded with enthusiastic applause. The bulk of Greene's talk was pretty much a summary of the work to date at the three IBM labs. But at the end of his talk, as a tease following the many reports of 'unidentified superconducting objects' (USOs) that were even then beginning to fly about, he jokingly reported that a minor insulating copper-oxide phase, called '211' (after its Y/Ba/Cu cation ratio), which precipitates during the synthesis of the 91-K superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$, or '123', had a transition temperature of 300 K — but as a ferroelectric, not a superconductor! This was a complete fabrication, but was swallowed by many. To this day, I'm occasionally informed by one of the unsuspecting, "Did you know that 211 is ferroelectric? Probably that's why it's such an effective pinning impurity in 123."

The happening

I was not present at 7 p.m. on Wednesday, when the doors of the Hilton's Grand Ballroom swung open to accept the surge of humanity — well, physicists — that then ensued. I was still having dinner with Müller, discussing some concerns he had over the unit-cell model for 123 that Robbie Beyers and Grace Lim in our group at IBM Almaden had determined from transmission electron microscopy and X-ray diffraction. Fortunately, the hotel had arranged reserved seating for the speakers, so on arriving at the ballroom we were able to make ourselves comfortable after manoeuvring through the swarm encamped on the floor and in the aisles. The talks started late, because of the difficulty of clearing this fractious mob — an issue that was finally resolved by the announcement that the hotel would provide television monitors in various lobbies that would carry the proceedings live.

In the end, 1,800 people squeezed into a room meant for 1,100, with 2,000 more outside watching the monitors.

Things began with a 'plenary' session of 15-minute talks given by Müller, Chu, Shoji Tanaka (University of Tokyo), Bertram Batlogg (AT&T Bell Laboratories) and Z. X. Zhao (Institute of Physics, Beijing). Much to my satisfaction, Müller showed as his last viewgraph our Almaden 123 unit cell, so dropping the first bombshell of the night. Another came shortly after from Batlogg, who waved, to tumultuous acclamation, a piece of '123 green sheet' (a flexible substrate with as-yet unreacted precursor oxides). Batlogg's showmanship implied that applications were not far away, and set the scene for much of the press attention that we were to receive in the coming months.

Around 3 a.m., even the faithful were beginning to tire, and many drifted into the lobbies for individual discussion and gossip, always with one eye on the video monitors in case some of the remaining speakers might reveal one more item not previously covered. I joined Ted Geballe of Stanford University, who was pouring champagne for all gathered around him. Geballe could well claim to have been the mentor of the generation that had brought about the post-Bednorz-Müller discoveries, and he was clearly in his element. Around daybreak, Engler and I wandered a few blocks over to a rather rundown saloon on 7th Avenue for a final nightcap and wind-down of our eight-hour adrenalin rush. Our visit was reported several days later in the *Wall Street Journal* as, "Grant and Engler staggered into a midtown bar..." much to the consternation, we later learned, of an IBM corporate vice-president when he read about it.

All of a sudden we physicists had become darlings of the media and the public. On Thursday night one of the most fashionable of New York discos, The Limelight, housed in a turn-of-the-century former Episcopal church on 20th Street, held a "physicists' night out". All one had to do was show up

with a meeting badge, to go to the front of the line, and in for free. The scene of several hundred physicists bopping to the strains of 'heavy metal' music was too much, and certainly not in accord with our previous public image. In the months to come, there would be many interviews and television appearances. In May, the three main US news magazines all had cover stories on high-temperature superconductivity. Wow, what a time we perceived antisocial PhD nerds were having. But the central question would soon become, "Is this all there is? Will advancement of science and beneficial social application emerge, despite the hype?"

Sobering up

The years following the original "Woodstock" had dissolved into the disillusionment of the Vietnam War. The plaintive lyrics of Pete Seeger's classic, "Where have all the flowers gone?", revived in the 1970s, well reflected the frustration of the era that followed. Was a similar fate to be the legacy of our Woodstock? Would the expectations sown by us in New York in the spring of 1987 ever blossom? The answer is, yes, some shoots are appearing today, but it has taken longer than we naively believed at the time.

The large number of scientists from almost all materials-related disciplines who poured into high-temperature superconductivity following the New York meeting was testimony to the enormous excitement that a discovery both mysterious and with promise of great application can promote. Of course, the theoreticians are the first to jump on anything, so their presence was no surprise. The extraordinary number of others who entered the field, many of them physicists and engineers who had previously had nothing to do with superconductivity, I believe has a lot to do with the ease with which one could fabricate $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. One of my most cited 'papers'³ concerned the success of a high-school chemistry class in California who had made and tested their own sample of 123 that May, only four months after its discovery, and five months before the award of the Nobel prize to Bednorz and Müller for starting it all. For a while you could attract almost any audience, including the President of the United States and other heads of state, with parlour-trick-like demonstrations of magnetic levitation.

There was also some progress of a substantial nature. By the time of the following March meeting, a number of new layered copper-oxide perovskites had been discovered (some might say rediscovered). One of them, containing thallium oxide sheets, set a world-record transition temperature (T_c) of 125 K (ref. 4), which stood until 1993. (A related structure, with HgO essentially supplanting the TlO portion⁵, now holds the ambient-pressure world record of 133 K.)

Nevertheless, a certain introspection was

beginning to surface. The 1988 March meeting hosted a symposium entitled "Physics and the press". A panel of journalists and physicists (of whom I was one) confessed a collective *mea culpa* over the excesses that each had committed: there were no levitated trains, no electricity too cheap to meter, no dime-sized supercomputers remotely on the horizon. I believe that the experience helped to moderate and make significantly more critical (with a few notable exceptions) the press response to the 1989 announcement of 'cold fusion'. In 1995 and 1996, high- T_c superconductivity began to make news again as power applications came closer. By and large, the coverage has not been sensationalized—I guess we learned our lesson.

Reality set in with a vengeance in late 1988 and throughout 1989–90. The high- T_c materials are type-II superconductors; that is, they allow an applied magnetic field to penetrate in a lattice of discrete 'vortex lines', the presence of which dominates almost all of the applied physics of superconductors. A great richness has arisen from the study of flux lattice dynamics in these very anisotropic systems—the resulting phase diagram has even been termed 'vortex matter'. One can arguably claim that the layered copper-oxide perovskites represent the very essence of type-II superconductivity. The art of useful applications lies in devising means to 'pin' the vortex lattice against dissipative flow arising from the Lorentz force imposed by a current or an external magnetic field. Working against the pinning potential are thermal fluctuations, which free the vortices to move—a phenomenon dubbed 'flux creep'.



Early showmanship: Bob Cava, of AT&T Bell Labs, displays for the press the 'green sheet' that his colleague Batlogg had waved in the evening session, along with a disk of the 123 superconductor. Now the superconductor itself can be made in flexible form.

In 1988, Michael Tinkham of Harvard University speculated that, because high- T_c superconductors would indeed be operated at relatively high temperatures, the vortex lattice might never remain completely pinned against flux flow, and therefore would not transport current without loss, particularly in high magnetic fields⁶. By this reasoning, if room-temperature superconductors were ever discovered, matters might only get worse.

On top of this, it was becoming clear that something beyond "shake 'n' bake" synthesis, so popular with us physicists, would be needed to turn these brittle and complex oxides, cousins to a teacup, into wires and electronic devices. Five, often six, elements were required to make some of the most promising layered copper oxides. No one had yet attempted to bring such complex materials into practical form. Many secondary phases would precipitate during the synthesis, resulting in serious degradation of the current-carrying capacity; in addition, difficulties associated with crystalline anisotropy seemed to be insurmountable. A news story in *Science* entitled "Superconductivity: is the party over?"⁷, which dwelt on the original promises of March 1987, and the likelihood of their remaining unfulfilled, was widely noted in press, industrial and government circles. Fortunately, the party was not yet over—the band was just out on a break.

Growing up

As the 1990s began, many of the less committed workers and institutions left the high- T_c field. But others, mainly materials scientists and new companies intent on commercialization, replaced the departed comrades and started to focus on critical problems, both fundamental and applied. In the early 1990s, the band returned from its break, and the party began to swing again, albeit to a more mature beat. It was as if Nature, having chastized us for our undue haste and optimism, finally lifted her penance, bestowing in addition some remarkable gifts.

Excellent and patient efforts by materials scientists worldwide led to better understanding of phase equilibria, especially in the 'BSCCO' family, comprising $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ (Bi-2212) and $\text{Bi}_2\text{Sr}_2\text{Ca}_2\text{Cu}_3\text{O}_{10}$ (Bi-2223). The substitution of small amounts of lead for bismuth was found to stabilize the 2223 phase, resulting in a bulk superconductor with a transition temperature of 110 K. Moreover, owing to the micaceous character of the BiO layers, both materials were far less brittle than might have been expected, which helped enormously in the drawing and rolling operations employed in wire manufacturing.

Perhaps even more important was the role of silver in inducing texturization in BSCCO, a process still incompletely understood. Most of the current carried by silver-sheathed BSCCO is transported by a layer a

few tens of micrometres thick, adjacent to the interface. Even so, as I write, the critical current density in metre-scale lengths exceeds $55,000 \text{ A cm}^{-2}$ at 77 K in self-field, and its sustainability at longer lengths is getting better and better.

Today, oxide-powder-in-silver-tube Bi-2223 and silver ribbon dip-coated with Bi-2212 are the mainstay products of 'Generation I' high-temperature superconducting wire in multi-kilometre lengths manufactured in the United States and Japan, and, after a late start, in Europe as well.

This wire is eminently suitable for electric power cables at 77 K, and for electromagnets, both stand-alone and in transformers and rotating machinery, at 30 K. Last year saw the public announcement of prototypes for both types of application⁸, and I can predict with confidence that the next two years will see their use in industry under engineering test conditions.

Meanwhile, flux creep has not turned out to be the problem it was once thought to be. Although Generation I wire does exhibit flux creep at 77 K, at the currents and magnetic fields typical of cable applications the resultant losses are much less than in conventional metals. The same is true at 30 K in magnetic fields of several tesla. At 4.2 K (liquid-helium temperature), high- T_c superconductors can operate in persistent-current (lossless) mode up to magnetic field strengths much higher than can niobium-titanium, the present mainstay technology of superconducting magnets in high-energy particle colliders and tokomaks.

And there is better news coming. New methods of processing good-old YBCO 123 in wire form are under development^{8,9}, taking advantage of its superior performance in high magnetic field and high temperature compared to BSCCO. It is reasonable to expect this 'Generation II' technology to mature as the millennium approaches. As for electronic applications, steady progress has been made in the development of both SQUID magnetometers and radio-frequency filters, and several companies are on the verge of introducing both passive and active thin-film devices as communications and sensor products⁹.

On the theoretical front, it did not long escape notice that high-temperature superconductivity in these systems invariably occurred in an insulating host crystal composed of two-dimensional sheets of antiferromagnetically coupled copper ions, and that the onset of superconductivity on doping to metallic levels of conductivity was probably not just an unrelated accident. A common feature of the 'magnetically mediated' models was the suspicion that the pairing symmetry was '*d-wave*' in nature, as opposed to the '*s-wave*' character found in phonon-coupled low- T_c materials. (It is important to note that 'plain vanilla' BCS

theory can accommodate both of these symmetries, and others as well.) So, around 1992, the attention of most experimentalists wishing to concentrate on 'mechanism-determining' measurements shifted to techniques designed to uncover the symmetry of the condensate wavefunction — less difficult, it was hoped, than a frontal attack on the pairing physics itself.

Improvements in the materials science of the layered copper-oxide perovskites helped provide the high-quality thin films and single crystals that were needed to probe the symmetry of the wavefunction. Most of these experiments suggest that the condensate state has *d*-wave symmetry, but data from direct tunnelling into both crystals and films of 123 require some *s*-wave component as well. To add to the confusion, measurements of penetration depth and thermomagnetotransport on the electron-doped compounds based on Nd_2CuO_4 imply *s*-wave symmetry. As yet, phase-sensitive experiments such as those done on 123 have not been possible in this latter system.

Sometimes I am asked by my applications-oriented colleagues whether it is really important to know what the pairing physics is in detail. The answer is, "Of course! We physicists are driven to know the how of things, and we would draw immense intellectual satisfaction from solving this puzzle." But would the answer aid in applications or serve as a guide to the discovery of new materials? Probably not.

Ten years young

Well, there you have it. Ten years later, over 50,000 papers published, more than a hundred compounds discovered, some elegant science advanced, but with the mechanism of high- T_c superconductivity as yet unresolved, and applications due in the next two years — the excitement of Woodstock continues. Scientific interest remains high, attested to by the suprisingly constant number of abstracts received by the newsletter *High- T_c Update* since its beginnings in 1987, and the frequent publication of papers in scientific journals of broad readership such as *Physical Review Letters*, *Nature* and *Science*.

On the other hand, several important corporations, such as IBM, AT&T and Bellcore — vital participants in the discovery years — are no longer major players. Their place has been taken by companies, both small and large, whose core businesses are much closer to the power and communications arenas, the historically natural targets for superconductivity. Interestingly, the best bulk and thin-film samples available for fundamental investigations are now made by these companies.

Government support for superconductivity remains high, hovering at about \$200 million per year in Japan, heavily oriented towards power, and \$150 million per year in

the United States, targeted mostly to electronics. The US national laboratories and Japan's ISTE, which has recently been refunded for another ten years, have been crucial partners with private industry, especially in power-related developments.

President Bill Clinton's 1998 budget submission includes a 60 per cent increase in the Department of Energy's superconductivity power programme, and, arguments about 'corporate welfare' aside, I think there is a good chance that Congress will approve it. Superconductivity has enjoyed bipartisan support in the past, and the advances wrought under the department's Superconductivity Partnership Initiative with industry are now vindicating that faith. Another very encouraging development has been the rapid growth in the past two years of involvement by European industry, utilities and government in power areas of superconductivity, now estimated at \$100 million per year.

One fallout from the large human and financial resources expended in the past ten years has been a level of worldwide patent activity in high-temperature superconductors rivalled only by semiconductor technology. Yet some fundamental patents have yet to be issued, primarily delayed by claims adjudication and interference proceedings. Who 'owns' high- T_c superconductivity remains to be determined. But one thing is certain: as applications emerge, there will be a lot of corporate lawyers knocking on each other's doors. You may want to follow my strategy in this regard. My wife Maria and I have two quite argumentative teenagers in our household (sound familiar?), whom we are encouraging to take their talents to law school and beyond, specializing in patent litigation. If high-temperature superconductivity ever produces substantial revenue returns, their parents' senior years will be secure. As a current Americanism for certainty puts it, "You can double-click on that icon."

Legacy for the next generation

What might our situation be ten years from now? Although another breakthrough such as we had in 1986–87 would be lovely, it is unlikely. Most probably, the layered copper-oxide perovskites are as good as we are going to get for some time. The current record transition temperature — held, quite appropriately, by Chu's institute in Texas — sits at 164 K in Hg-1223 at high pressure¹⁰, establishing an existence proof that at least this level might be possible in some unknown ambient-pressure phase. But my own instincts are that the copper oxides have already been played out. No other material system, not even semiconductors, has received such intense scrutiny. If there were more gold to be mined, it would have been struck by now, given the talent and tools dedicated to the search over the past ten years.

Since Woodstock, three other material systems that might have qualified in the 'old days' for the appellation "high- T_c " have been found. These are the cubic perovskite $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ (ref. 11), the alkali-metal-doped fullerenes¹² and the Pd and Ni borocarbides¹³, with transition temperatures ranging from 15 to 30 K. None holds strong promise for application because of their relatively low (by present standards!) values of T_c .

The next breakthrough in superconducting materials will come, as it always has in the past, from an unexpected source, maybe in organic or biological compounds. Perhaps the discovery will not even involve superconductivity, but rather 'perfect conductivity' of an exotic nature, mediated by a soliton or charge density wave mechanism, as has been proposed from time to time. In the meantime, one important legacy of the copper-oxide perovskites, beyond giving the world high-temperature superconductivity, is the renaissance now occurring in the transition-metal oxide family, typified by the detection of new effects such as giant and colossal magnetoresistance.

It might surprise some that many of the veterans of Woodstock are still actively involved in high- T_c superconductivity. It shouldn't — most of us were working on superconductivity well before 1986, and the phenomenon is incurably addictive. Significant numbers of people, myself included, are no longer with our former institutions, having moved on to universities or to government and private agencies with a focus on superconductivity, or into new commercial ventures to capitalize on the Bednorz–Müller discovery. Over the years, despite our past and present competitive battles, we have come to form, like Henry with his nobles and archers at Agincourt, a "band of brothers", whom others envy because we were there.

Sadly, there will be no official recognition of the ten years since Woodstock at this year's March meeting in Kansas City. Perhaps this is as it should be; New York was the quintessential stage on which to play out that drama, and to celebrate anywhere else would be inappropriate. Nevertheless, a few of us old soldiers will most surely gather in some mid-west waterhole to raise a cup in remembrance of 'the way it was'. Cheers. □

Paul M. Grant is at the Electric Power Research Institute, PO Box 10412, 3412 Hillview Avenue, Palo Alto, California 94303, USA.

1. Wu, M. K. et al. *Phys. Rev. Lett.* **58**, 908–910 (1987).
2. Bednorz, J. G. & Müller, K. A. *Z. Phys. B* **64**, 189–193 (1986).
3. Grant, P. M. *New Scientist* **115**, 36–39 (1987).
4. Parkin, S. S. P. et al. *Phys. Rev. Lett.* **60**, 2539–2542 (1988).
5. Schilling, A., Cantoni, M., Guo, J. D. & Ott, H. R. *Nature* **363**, 56–58 (1993).
6. Tinkham, M. *Phys. Rev. Lett.* **61**, 1658–1661 (1988).
7. Pool, R. *Science* **244**, 914–916 (1989).
8. Grant, P. M. *Nature* **381**, 559–560 (1996).
9. Grant, P. M. *Nature* **375**, 107–108 (1995).
10. Gao, L. et al. *Phys. Rev. B* **50**, 4260–4263 (1994).
11. Cava, R. J. et al. *Nature* **332**, 814–816 (1988).
12. Hebard, A. F. et al. *Nature* **350**, 600–601 (1991).
13. Cava, R. J. et al. *Nature* **367**, 252–253 (1994).

Clone mammals... clone man?

What are the implications for humankind of the astounding report two weeks ago of the production of viable sheep from adult cells? The moral imperative of preserving human dignity must remain paramount.

Axel Kahn

The experiments of I. Wilmut *et al.* (*Nature* 385, 810; 1997) demonstrate that sheep embryonic eggs (oocytes) can reprogramme the nuclei of differentiated cells, enabling the cells to develop into any type. The precise conditions under which this process can occur remain to be elucidated; the factors determining the success of the technique, and the long-term prospects for animals generated in this way, still need to be established. But, of course, the main point is that Wilmut *et al.* show that it is now possible to envisage cloning of adult mammals in a completely asexual fashion.

The oocyte's only involvement is the role of its cytoplasm in reprogramming the introduced nucleus and in contributing intracellular organelles — mainly mitochondria — to the future organism. This work will undoubtedly open up new perspectives in research in biology and development, for example, in understanding the functional plasticity of the genome and chromatin during development, and the mechanisms underlying the stability of differentiated states. Another immediate question is to ask whether a species barrier exists. Could an embryo be produced, for example, by implanting the nucleus of a lamb in an enucleated mouse oocyte? Any lambs born in this way would possess a mouse mitochondrial genome.

The implications for humans are staggering. One example is that the technique suggests that a woman suffering from a serious mitochondrial disease might in future be able to produce children free of the disease by having the nucleus of her embryo implanted in a donor oocyte (note that this process is not the same as 'cloning').

Would cloning humans be justified?

But the main question raised by the paper by Wilmut *et al.* is that of the possibility of human cloning. There is no *a priori* reason why humans should behave very differently from other mammals where cloning is possible, so the cloning of an adult human could become feasible using the techniques reported.

What medical and scientific 'justification' might there be for cloning? Previous debates have identified the preparation of immunocompatible differentiated cell lines for transplantation, as one potential indication. We could imagine everyone having their own reserve of therapeutic cells that would increase their chance of being cured of various diseases, such as cancer, degenerative disorders and viral or inflammatory diseases.

But the debate has in the past perhaps paid

insufficient attention to the current strong social trend towards a fanatical desire for individuals not simply to have children but to ensure that these children also carry their genes. Achieving such biological descent was impossible for sterile men until the development of ICSI (intracytoplasmic sperm injection), which allows a single sperm to be injected directly into the oocyte.

But human descent is not only biological, as it is in all other species, but is also emotional and cultural. The latter is of such importance that methods of inheritance where both parents' genes are not transmitted — such as adoption and insemination with donor sperm — are widely accepted without any major ethical questions being raised.

But today's society is characterized by an increasing demand for biological inheritance, as if this were the only desirable form of inheritance. Regrettably, a person's personality is increasingly perceived as being largely determined by his or her genes. Moreover, in a world where culture is increasingly internationalized and homogenized, people may ask themselves whether they have anything else to transmit to their children apart from their genes. This pressure probably accounts for the wide social acceptance of ICSI, a technique which was widely made available to people at a time when experimental evidence as to its safety was still flimsy. ICSI means that men with abnormal sperm can now procreate.

Going further upstream, researchers have now succeeded in fertilizing a mouse oocyte using a diploid nucleus of a spermatogonium: apparently normal embryonic development occurs, at least in the early stages. But there are also severe forms of sterility — such as dysplasia or severe testicular atrophy — or indeed lesbian couples, where no male germ line exists. Will such couples also demand the right to a biological descent?

Applying the technique used by Wilmut *et al.* in sheep directly to humans would yield a clone 'of the father' and not a shared descendant of both the father and mother. Nevertheless, for a woman the act of carrying a fetus can be as important as being its biological mother. The extraordinary power of such 'maternal reappropriation' of the embryo can be seen from the strong demand for pregnancies in post-menopausal women, and for embryo and oocyte donations to circumvent female sterility. Moreover, if cloning techniques were ever to be used, the mother would be contributing something — her mitochondrial genome. We cannot exclude the possibility that the current direction of public opinion will tend to legitimize the resort to cloning techniques in cases

where, for example, the male partner in a couple is unable to produce gametes.

The creation of human clones solely for spare cell lines would, from a philosophical point of view, be in obvious contradiction to an ethical principle expressed by Emmanuel Kant: that of human dignity. This principle demands that an individual — and I would extend this to read human life — should never be thought of only as a means, but always also as an end. Creating human life for the sole purpose of preparing therapeutic material would clearly not be for the dignity of the life created.

Analysing the use of cloning as a means of combating sterility is much more difficult, as the explicit goal is to create a life with the right to dignity. Moreover, individuals are not determined entirely by their genome, as of course the family, cultural and social environment have a powerful 'humanizing' influence on the construction of a personality. Two human clones born decades apart would be much more different psychologically than identical twins raised in the same family.

Threat of human 'creators'

Nonetheless, part of the individuality and dignity of a person probably lies in the uniqueness and unpredictability of his or her development. As a result, the uncertainty of the great lottery of heredity constitutes the principal protection against biological predetermination imposed by third parties, including parents. One blessing of the relationship between parents and children is their inevitable difference, which results in parents loving their children for what they are, rather than trying to make them what they want. Allowing cloning to circumvent sterility would lead to it being tolerated in cases where it was imposed, for example, by authorities. What would the world be like if we accepted that human 'creators' could assume the right to generate creatures in their own likeness, beings whose very biological characteristics would be subjugated to an outside will?

The results of Wilmut *et al.* undoubtedly have much merit. One effect of them is to oblige us to face up to our responsibilities. It is not a technical barrier that will protect us from the perspectives I have mentioned, but a moral one, originating from a reflection of the basis of our dignity. That barrier is certainly the most dignified aspect of human genius. □

Axel Kahn is director of the INSERM Laboratory of Research on Genetics and Molecular Pathology at the Cochin Institute of Molecular Genetics, 75014 Paris, France.

This article is a slightly edited version of a Commentary first published on Nature's Web site on 27 February.

True vision of a quantum state

Matthias Freyberger and Wolfgang P. Schleich

Information about a quantum system is encoded in its quantum state, a quantity whose meaning is vigorously debated. But direct insight should be gained into quantum states now that they can be mapped out.

Physicists work hard to unravel the information contained in systems that they encounter in nature. Of course, they are not just collecting data to set up a catalogue; primarily, they are interested in finding relationships that enable them to predict new phenomena.

Quantum mechanics formulates these relationships in the microscopic quantum world — a world where light behaves as particles and matter behaves as waves. The most fundamental quantity of quantum mechanics is the quantum state, which encodes all the information about a system. Hence a direct measurement of this state is one of the great challenges of experimental physics¹.

A few state measurements of quantum systems have already been made: a quantized light field^{2,3}, a vibrating molecule⁴ and an atom in a Paul trap⁵. On page 150 of this issue⁶, Kurtsiefer, Pfau and Mlynek now report a further important step towards demystifying the concept of a quantum state. They have succeeded in measuring the quantum state of atoms impinging on a double slit.

According to classical physics, the position x and momentum p of an atom define its state. The atom moves like a little ball that either hits the screen or passes through one of the two slits. Quantum mechanics changes this picture drastically. Heisenberg's uncertainty relation tells us that we cannot know — as a matter of principle — the position and momentum of a particle simultaneously with arbitrary precision; so the classical definition of a state breaks down. Nevertheless, quantum theory sticks to the concept of a state ψ that can be represented as a complex-valued function of position. This function $\psi(x)$, the so-called wavefunction, contains all the information about a quantum particle. In particular, it allows us to determine the probability $|\psi(x)|^2$ of detecting the particle at position x .

So according to quantum theory, the atoms travel through the double slit not like classical particles but like waves, and each atom has a specific probability of passing through one slit or the other. Parts of these probability waves passing through one slit interfere with those that come from the other slit, and Kurtsiefer *et al.* see this interference (Fig. 2, page 151).

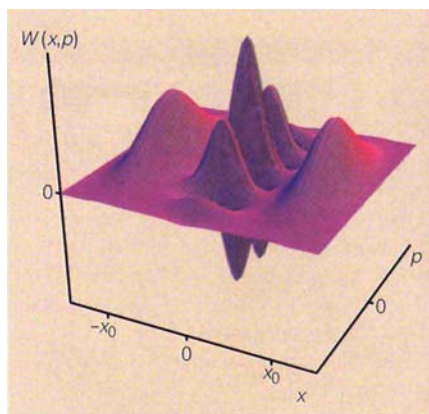


Figure 1 The Wigner function of an atomic matter wave after passing through a double slit with apertures at x_0 and $-x_0$. The strong oscillations leading to negativities express the non-classical wave nature of atoms passing through the double slit.

But they also reconstruct the complete atomic quantum state that lies at the heart of this matter-wave interference. Instead of representing this state as a wavefunction, they use a more vivid presentation of it — the Wigner function $W(x, p)$. The Wigner function contains the same information as the wavefunction, but it has the advantage of being a real valued function, which can be illustrated in a phase space spanned by position and momentum. However, it is not a joint probability distribution of position and momentum, because that would contradict the uncertainty relation.

A peculiarity of the Wigner function is that it can take on negative values, the signatures of non-classical behaviour. Figure 1 shows a theoretically calculated Wigner function for the quantum state of a matter wave of helium atoms after it has passed through a double slit. The crests and troughs signify the non-classical behaviour imprinted on the atomic motion by the slit.

Kurtsiefer *et al.*⁶ report the first experimental evidence for such non-classical signatures in a matter wave by measuring this Wigner function. This is, of course, not a straightforward experimental task, because it is impossible to detect in a classical sense: we cannot obtain the atomic statistics by just measuring position and momentum of the atoms and by counting how often they occur

— that is forbidden by the Heisenberg uncertainty relation. So we have to adopt more sophisticated strategies.

Kurtsiefer *et al.* used the phase-space tomography method^{7,8}, measuring the positions and arrival times of atoms from the helium beam. The information encoded in a quantum state is also stored in a specific set of probability distributions for the atomic positions, each distribution being a projection of the Wigner function on the position axis. Hence, a single position distribution does not contain enough information to reconstruct the quantum state. This would be like reconstructing a mountain from just one specific view given, for example, on a postcard. But if we collect many postcards showing different views of the same mountain, then it might be possible to build a three-dimensional model; equivalently, we collect a set of position distributions obtained by projecting rotated versions of the 'Wigner mountain' onto the x axis. The symmetry of the Wigner function determines how many directions we have to look at before there are enough data for a reconstruction; in general, phase-space tomography needs position distributions for every rotation angle between 0° and 180° . What remains is a tedious mathematical transformation — the inverse Radon transform — which allows us to convert this set of position distributions into the original Wigner function.

What we simply called a rotation has to be implemented as a physical process. Fortunately, the Wigner function rotates during free propagation of the atoms⁹, and the rotation angle then depends on the time of flight from the slits to the detection screen. From an experimental point of view this is the most elegant and direct approach. But free propagation cannot lead to a complete rotation of the original Wigner function over a full 180° interval¹⁰, so Kurtsiefer *et al.* have to rely on specific symmetries of the Wigner function they are going to measure.

In their case, this is possible because the physical object, namely the double slit preparing the Wigner function, is simple. In the general case, however, free propagation is not enough. Instead, one can use the interaction of the atom with a harmonic potential. The time evolution of the atom in the potential then allows for arbitrary rotations of a Wigner function. A drawback of this method, which makes it more difficult experimentally, is the realization of the harmonic potential. One possibility is to send the atoms through a standing light field, which acts like a harmonic potential as long as its wavelength is larger than the spatial extent of the Wigner function¹¹.

Quantum state measurements open a new and fascinating road into the microscopic world of quantum physics. We are far

from the end of this road, and many more ideas¹² about quantum-state measurements are already on the horizon. The work of Kurt-Siefer *et al.* takes an important step in the direction of unravelling the most fundamental quantity of quantum theory. □

Matthias Freyberger and Wolfgang P. Schleich are in the Department of Quantum Physics, University of Ulm, D-89069 Ulm, Germany.

- Schleich, W. P. & Raymer, M. G. (eds) *J. Mod. Opt. Spec. issue* (in the press).
- Smithey, D. T., Beck, M., Raymer, M. G. & Faridani, A. *Phys. Rev. Lett.* **70**, 1244–1247 (1993).
- Schiller, S., Breitenbach, G., Pereira, S. F., Müller, T. & Mlynek,

J. Phys. Rev. Lett. **77**, 2933–2936 (1996).

- Dunn, T. J., Walmsley, I. A. & Mukamel, S. *Phys. Rev. Lett.* **74**, 884–887 (1995).
- Leibfried, D. *et al. Phys. Rev. Lett.* **77**, 4281–4285 (1996).
- Kurtsiefer, Ch., Pfau, T. & Mlynek, J. *Nature* **386**, 150–153 (1997).
- Bertrand, J. & Bertrand, P. *Found. Phys.* **17**, 397–405 (1987).
- Vogel, K. & Risken, H. *Phys. Rev. A* **40**, 2847–2849 (1989).
- Janicke, U. & Wilkens, M. *J. Mod. Opt.* **42**, 2183–2199 (1995).
- Raymer, M. G. in *J. Mod. Opt. Spec. issue* (eds Schleich, W. P. & Raymer, M. G.) (in the press).
- Kienle, S., Freyberger, M., Schleich, W. P. & Raymer, M. G. in *Experimental Metaphysics: Quantum Mechanical Studies for Abner Shimony* (eds Cohen, R. S., Horne, M. A. & Stachel, J. S.) (Kluwer, Dordrecht, in the press).
- Kokorowski, D. A. & Pritchard, D. E. in *J. Mod. Opt. Spec. issue* (eds Schleich, W. P. & Raymer, M. G.) (in the press).

Neurobiology

Look but don't touch, or vice versa

Michael Shadlen

Nowhere in the brain is the connection between body and mind so conspicuous as in the parietal lobes: damage to the parietal cortex disrupts awareness of one's body and the space that it inhabits^{1,2}. Experiments in monkeys indicate that the neural circuits responsible for spatial perception comprise a pathway — the 'where' pathway — which leads to the posterior parietal cortex (PPC)³. However, the PPC also makes connections with the motor structures that are involved in planning and executing movements of the hands and eyes^{4–8}. So do neurons in the parietal lobe mainly act to represent the spatial whereabouts of objects, or do they guide parts of the body to the location that these objects occupy? The former view has led to the idea that the PPC mediates attention to locations in personal and extrapersonal space. But on

page 167 of this issue, Snyder, Batista and Andersen⁹ suggest that neurons in the PPC are instead concerned with specific motor functions. Rather than directing attention to objects and locations, Snyder *et al.* propose that the PPC signals an *intention* to do something.

In monkeys, neurons around the intraparietal sulcus respond to visual targets in an intriguing manner. They are activated by objects that appear within a restricted region of the visual field, but they do not seem to be too fussy about the nature of the object — any colour or pattern will do, as long as it falls in the right location. Moreover, the neural response can persist for several seconds after the target has disappeared, but only if the location of the target retains its significance for behaviour.

The response shown in Fig. 1a is typical of

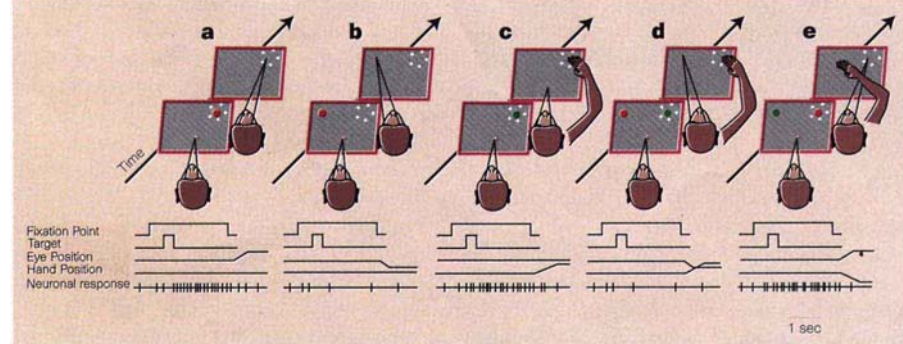


Figure 1 Simulated response of a neuron in the posterior parietal cortex (PPC) of a monkey to oculomotor and reaching tasks, as studied by Snyder *et al.*⁹. The response pattern would be typical of neurons located laterally in the PPC. Upper cartoons are snapshots of the task as viewed by the monkey; lower traces show the time course of the fixation point, eccentric target(s), idealized eye and hand positions, and response of the neuron. The dotted circular region represents the visual receptive field (or movement field) of the neuron — the circle is not visible to the monkey. a, The neuron responds when the monkey plans an eye movement to the location cued by the target: the 'go' signal is the extinction of the fixation point. b, There is no response when the target and eye movement are outside the appropriate location. c, Many parietal neurons also respond when the monkey plans a reaching movement to the cued location but, according to Snyder *et al.*, this is because the monkey covertly plans an eye movement. d, e, The monkey reaches for the location cued by the green target, and simultaneously shifts its gaze to the location cued by the red target. The neuron responds only when the monkey plans an eye movement to the upper right. In medial regions of the PPC, the opposite pattern of results is obtained.

this. If a target appears briefly in a region of the visual field outlined by the dashed circle, the neuron responds with a volley of activity that persists until the monkey shifts its gaze to the cued location. The same neuron does not respond when the target appears in a different location, or if the monkey is instructed to shift its gaze elsewhere (Fig. 1b). The response is therefore said to be spatially selective, and linked in some way to the behavioural contingencies of the task. Beyond this it is difficult to interpret the response; for example, whether the neuron codes for a location in space, or a plan to move the eyes there.

About fifteen years ago, Goldberg and colleagues¹⁰ examined this question by training monkeys to reach for a target in some trials and to shift their gaze to the target in others. They found many neurons that discharged when a monkey reached for a target, as shown in Fig. 1c. Interestingly, many neurons responded similarly to the cued location, whether the monkey made an eye movement or a hand movement there, supporting the hypothesis that the PPC encodes the location of objects or the focus of visual attention.

Snyder *et al.*⁹ added a new twist to this gaze-versus-reach task, with surprising results. They trained monkeys either to reach for, or to shift their gaze towards, a cued location, but they ensured that only one or the other action was done. In some trials, the monkey touched a button while holding its gaze fixed; in others, the monkey looked towards the target while holding its hand still. Under these conditions, most neurons in the lateral region of the PPC discharged before eye movements, but not before reaches. In a more medial region, the authors found neurons that responded when the monkey reached, but not when it made eye movements. They also found many neurons that responded to the cued location regardless of which action was solicited. They then taught the monkey a 'dissociation' task — to move its eyes to one location while reaching for another (Fig. 1d and e). Not surprisingly, those neurons that responded only before eye movements did so on this task as well, responding when the monkey planned an eye movement to the appropriate part of the visual field. Reach-selective neurons also responded predictably on this task.

So, what about the neurons that responded on either action? Most of these neurons were either of the 'gaze' or the 'reach' variety. Those in the more medial regions — in the vicinity of reach neurons — no longer fired when the monkey made an eye movement to the cued location, because a hand movement was planned elsewhere: these neurons only responded when the monkey planned to reach to the appropriate location. Neu-

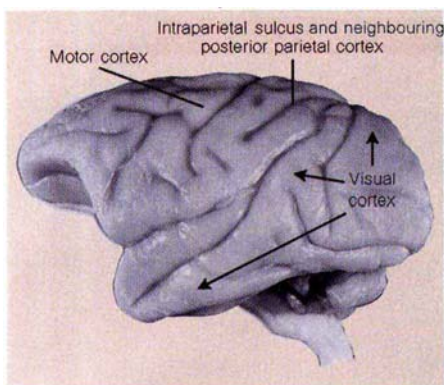


Figure 2 The brain of a rhesus monkey viewed from its left surface. The neurons studied by Snyder *et al.*⁹ are in the posterior parietal cortex, and this area is thought to be essential for coordinating spatial vision, attention and visually guided behaviour such as reaching.

rons in the more lateral region, where eye-movement-related neurons predominate, fired only when the monkey planned eye movements. This response pattern is shown in Fig. 1d and e.

Snyder *et al.*⁹ conclude that neurons in the PPC reflect the monkey's intention to do something in particular: they do not represent the location of objects in a manner that is detached from what the animal intends to do about it. Although this conclusion is tantalizing, it must be viewed with some caution. For instance, specificity for motor intention may be less evident in the early phase of the response, just after the monkey sees the visual target (see Fig. 1, page 167). Moreover, when the monkey was not required to dissociate reach and gaze, one-third of the neurons responded to the visual target, irrespective of whether the monkey planned to reach for or look towards its location. Snyder *et al.* interpret this lack of selectivity as a covert plan to perform either action — a strategy that is only precluded through the dissociation task. The idea is that the brain naturally queues up several motor plans, only a subset of which come to be executed. But is this really distinct from the concept of attention? After all, what is meant by attention to a location if not the possibility of directing one's gaze or hand there¹¹?

The study raises several other questions. Could the preponderance of neurons with specific allegiances to motor systems have emerged because of training on the dissociation task? Suppose the monkey had been trained to coordinate hand and eye movements to the same location: would the PPC contain more neurons that respond to both modes of behaviour? Perhaps a new class of neurons would emerge which respond selectively only when gaze and reach are coordinated; such neurons are present in the supplementary eye fields of the frontal lobe¹².

More fundamental questions concern the mechanism that underlies the parietal response; for example, how visual activation gives rise to a sustained response reflecting intention. Such conversion of visual sensation to a behavioural plan seems to be at the heart of cognition¹³. Where in the brain do spatial targets and instructions interact to produce a plan to look, reach, or look and reach? How do these intention-related signals ultimately affect the motor structures that execute the plan? And what happens to these neural signals when a covert plan is aborted?

It is intriguing that the 'where' pathway for vision should culminate in neural structures that are linked to specific body movements. This observation indicates that the brain organizes its information in a framework that is defined by its repertoire of motor functions (Fig. 2). So, for the brain, spatial location is not a mathematical abstraction or a property of a map, but it involves the issue of how the body navigates its hand or gaze^{14,15}. If the parietal lobe is indeed responsible for our appreciation of space — be it the position of objects or the arrangement of their parts —

then it is fascinating that this organization should reflect motor intention. □

Michael Shadlen is in the Department of Physiology and Biophysics and the Regional Primate Center, University of Washington School of Medicine, Seattle, Washington 98195-7290, USA.

1. Critchley, M. *The Parietal Lobes* (Hafner, New York, 1953).
2. Mesulam, M.-M. (ed.) in *Principles of Behavioral Neurology* 125–168 (Davis, Philadelphia, PA, 1985).
3. Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 549–580 (MIT Press, Cambridge, MA, 1982).
4. Hyvarinen, J. & Poranen, A. *Brain* **97**, 673–692 (1974).
5. Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H. & Acuna, C. *J. Neurophysiol.* **38**, 871–908 (1975).
6. Colby, C. & Duhamel, J.-R. *Neuropsychologia* **29**, 497–515 (1991).
7. Sakata, H., Taira, M., Murata, A. & Mine, S. *Cerebral Cortex* **5**, 429–438 (1995).
8. Dottie, M. C. *et al. Nature* **383**, 618–621 (1996).
9. Snyder, L. H., Batista, A. P. & Andersen, R. A. *Nature* **386**, 167–170 (1997).
10. Bushnell, M. C., Goldberg, M. E. & Robinson, D. L. *J. Neurophysiol.* **46**, 755–772 (1981).
11. Rizzolatti, G., Riggio, L., Dascola, I. & Umiltà, C. *J. Neuropsychologia* **25**, 31–40 (1987).
12. Mushiake, H., Fujii, N. & Tanji, J. *J. Neurophysiol.* **75**, 2187–2191 (1996).
13. Shadlen, M. N. & Newsome, W. T. *Proc. Natl Acad. Sci. USA* **93**, 628–633 (1996).
14. Merleau-Ponty, M. *Phenomenology of Perception* (Routledge & Kegan Paul, London, 1962).
15. Helmholtz, H. L. F. *Physiological Optics* (Dover, New York, 1925).

Quantum chaos

Fractal resistance in a transistor

Mark Fromhold

The recognition that even simple systems can have extremely complex behaviour has revolutionized many branches of science¹. In mathematics, simple formulae can generate beautiful images that bear an eerie resemblance to natural landscapes². These images contain self-similar or 'fractal' patterns, which repeat on smaller and smaller scales. Now Taylor *et al.*³ have seen remarkable fractal structure in the resistance of a small semiconductor device. The results raise fundamental questions about the relationship between classical chaos theory and quantum mechanics.

Newtonian classical mechanics applies to the large objects we encounter in everyday life. With the advent of computers came calculations revealing that simple dynamical systems obeying Newton's laws of motion often have highly erratic behaviour. The Sinai billiard⁴ has served as a model system for the study of chaotic classical motion since the early 1970s. The billiard table (Fig. 1) consists of a flat frictionless surface with a square (or rectangular) boundary enclosing an inner circular wall from which the ball bounces with no loss of energy. A billiards player would find this table frustrating because the ball follows a highly irregular path which depends critically on how it is initially struck. This chaotic behaviour is in marked contrast with the highly predictable

path followed by a ball on a conventional rectangular billiard table.

For small objects such as atoms, classical mechanics fails spectacularly. It cannot, for example, explain why an electron orbiting an atomic nucleus only has certain discrete energy levels. The properties of atoms can, however, be predicted to astonishing accuracy by quantum theory in which the electron is viewed as a wave rather than a charged ball. Interference of the electron waves determines the allowed or 'quantized' energy levels of the electron. Over the past 20 years,

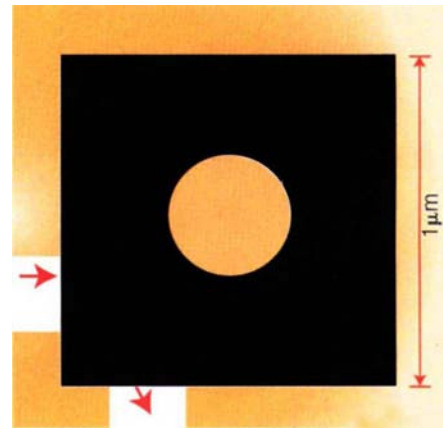


Figure 1 Chaotic resistor. The device studied by Taylor *et al.*³ is identical to the Sinai billiard⁴, except that the square gate contains two small holes to allow electrons to flow in and out.

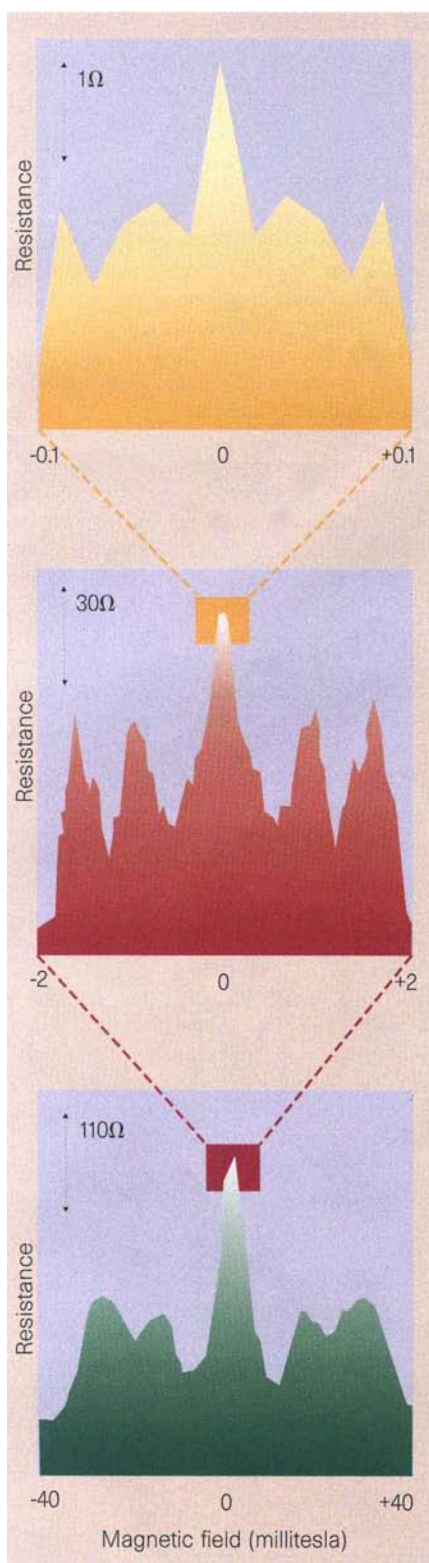


Figure 2 Fractal resistance. The resistance of the Sinai billiard is plotted as a function of magnetic field strength (field applied perpendicular to the plane of the device) at a temperature of 30 mK. The resistance fluctuations observed on the finest magnetic-field scale are approximately two orders of magnitude smaller than those seen in previous studies of quantum billiards. The fluctuation pattern repeats itself on magnification, but it isn't known how far the self-similarity continues.

quantum chaos — the search for the fingerprints of chaotic classical motion in the quantum world — has emerged as one of the most challenging topics in contemporary science⁵.

Advances in crystal growth techniques have led to the fabrication of a range of semiconductor devices that provide flexible tools for the study of quantum chaos^{6–9}. Collaborators at the Institute for Microstructural Sciences, National Research Council, Ottawa, the University of New South Wales, and McGill University³ have now exploited semiconductor technology to make a Sinai billiard small enough to show both chaotic and quantum-mechanical effects. The device is a semiconductor sandwich structure in which the conduction electrons are constrained to move as a two-dimensional electron gas. A metallic layer or 'gate' in the shape of a square with two small holes (Fig. 1) is deposited on the upper surface of the device. When a negative voltage is applied to this gate, the electrons below the gate are repelled from it.

This leads to the formation of a two-dimensional square box with electrostatic walls created by the electric forces underneath the gate. Electrons enter and leave the box through the openings in the boundary wall and, in an ideal square box, move along predictable classical paths. To generate chaotic classical orbits, Taylor *et al.*³ also deposited a circular surface gate which is electrically insulated from the surrounding square gate. When a negative voltage is applied to the circular gate, the device is transformed from a square into an electrical analogue of the famous Sinai billiard. This ability to make a controlled transition between stable and chaotic electron orbits in a single billiard structure is a considerable advance. Previously it has only been possible to compare the properties of chaotic and non-chaotic billiards by using two different devices.

A small voltage V between the entrance and exit holes drives a current I between them, and the electrical resistance R of the device is simply V/I . Information about the motion of electrons in quantum billiards is commonly obtained by seeing how the resistance of the device changes when a magnetic field is applied perpendicular to the sheet of electrons. In classical mechanics, the magnetic field bends the electron path into a series of circular arcs. But in quantum mechanics it produces complex changes in the interference patterns of the electron waves flowing through the device.

Quantum interference has already been shown to generate rich fluctuation patterns in the resistance versus magnetic field plots for a variety of semiconductor billiards^{6,7,9}. But the experiments of Taylor *et al.* are of particular interest because they provide the first observation of self-similarity (fractal-

like patterns) in the resistance of a semiconductor. This self-similarity is easily seen by comparing the shapes of the curves shown for three different magnetic field scales in Fig. 2. Each successive magnification of the largest resistance peak reproduces the same pattern of fluctuations. The pattern is fractal in the sense that it repeats on three distinct magnetic field scales that are related by a constant magnification factor of 20. Adjusting the gate voltages changes the resistance pattern in the same way for each level of magnification.

The physical origin of these fractal resistance fluctuations is not understood. An important clue could be that the material quality of the new device is so high that the electrons travel much further before scattering off structural imperfections than in any older semiconductor billiard. As a consequence, the resistance of the device is sensitive to quantum interference effects associated with long classical orbits in which the electron bounces around the billiard many times before exiting. Theoretical modelling of these interference effects may help to explain why the fractal resistance fluctuations are even seen on an extremely small magnetic field scale of around 50 microtesla. The electron gas is also further from the surface gates than in earlier samples. This produces 'soft' confining walls which have been linked to fractal resistance fluctuations originating from self-similar sets of classical orbits in small billiards¹⁰ and gold nanowires¹¹.

Further work is needed to determine whether fractal resistance fluctuations are a generic feature of chaotic quantum billiards or a peculiarity of the Sinai geometry, and to explore the tantalizing possibility that the fractal patterns repeat on even finer scales beyond the resolution of the current experiments. It is not clear whether this research will have any impact on the development of semiconductor microchips. Nevertheless, the results provide a most welcome link between computer simulations of the quantum mechanical properties of chaotic billiards and the real physical world of electronic motion through micro-transistors. □

Mark Fromhold is in the Department of Physics, University of Nottingham, Nottingham NG7 2RD, UK.

1. Stewart, I. *Does God Play Dice? The Mathematics of Chaos* (Penguin, London, 1990).
2. Mandelbrot, B. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
3. Taylor, R. P. *et al.* *Phys. Rev. Lett.* **78**, 1952–1955 (1997).
4. Sinai, Ya. G. *Russ. Math. Survey.* **25**, 137–189 (1970).
5. Berry, M. V. *Proc. R. Soc. Lond. A* **413**, 183–198 (1987).
6. Marcus, C. M., Rimberg, A. J., Westervelt, R. M., Hopkins, P. F. & Gossard, A. C. *Phys. Rev. Lett.* **69**, 506–509 (1992).
7. Chang, A. M., Baranger, H. U., Pfeiffer, L. N. & West, K. W. *Phys. Rev. Lett.* **73**, 2111–2114 (1994).
8. Wilkinson, P. B. *et al.* *Nature* **380**, 608–610 (1996).
9. Bird, J. P. *et al.* *Europhys. Lett.* **35**, 529–534 (1996).
10. Ketzmerick, R. *Phys. Rev. B* **54**, 10841–10844 (1996).
11. Heeger, H. *et al.* *Phys. Rev. Lett.* **77**, 3885–3888 (1996).

NATURE

100 YEARS AGO

Our congratulations to the Leicester Literary and Philosophical Society. Stimulated into action by a paper on the disappearances of certain species of insects, by Mr. Frank Bouskell, a Committee was formed to formulate regulations for the protection of local species... When a collector now sees *Leucophasia sinapsis* (the Wood White butterfly), he must hold his hand and crush his sporting instinct, for none of this insect are to be taken. Of *Macroglossa fuciformis* only one specimen must be taken by each member in a single season, and only one specimen of *Sesia apiformis*. The penalties for breaking these regulations are drastic... The over-zealous collector will, indeed, be ostracised, and will find that no one will buy from him, exchange with him, or have anything to do with him entomologically. There may be a difficulty in carrying out the regulations, and one result will probably be that collectors will prefer to go out alone in the future. But it is hoped that entomologists will remember that they are not supposed merely to fulfil the functions of a fly-paper, but also to work for the advancement of their science.

From *Nature* 11 March 1897.

50 YEARS AGO

When Prof. Raymond Dart, of the University of the Witwatersrand, Johannesburg, announced in *Nature* the discovery of a juvenile *Australopithecus* and claimed for it a human kinship, I was one of those who took the point of view that when the adult form was discovered it would prove to be near akin to the living African anthropoids. [But] I am now convinced that Prof. Dart was right and that I was wrong; the *Australopithecinae* are in or near the line which culminated in the human form. My only complaint now is the length of the name which the extinct anthropoid of South Africa must for ever bear... I have ventured, when writing of the *Australopithecinae*, to give them the colloquial name of 'Dartians', thereby saving much expenditure of ink and of print... It is much easier to say there was a 'Dartian' phase in man's evolution than to speak of one which was 'australopithecine'. — Arthur Keith

From *Nature* 15 March 1947.

Xenotransplantation

Provirus pose potential problems

Jonathan P. Stoye

Organ transplantation could save — or greatly improve the quality of — the lives of many people. However, there is a shortage of suitable donated human organs, so attention has turned to the possibility of transplanting primate or pig organs. Using transgenic animals designed to minimize the activation of human complement, considerable progress has been made in preventing immunological rejection of transplanted organs^{1,2}. Yet clinical trials have been delayed, mainly due to fears that such procedures might facilitate the introduction of novel viruses, particularly retroviruses, to the human population³. This was largely a theoretical risk, at least in the case of pigs, as no such viruses had been described. But a report by Patience and colleagues⁴ in the March issue of *Nature Medicine* indicates that there are grounds for practical concern. They provide evidence that at least one pig cell line produces a retrovirus that can infect human cells, and that a related virus is expressed in normal pig tissues.

If a retrovirus infects a germ cell, the resulting provirus can become part of the germ line. These retroviral elements are known as endogenous proviruses, and every vertebrate species will have picked up many such mementos of encounters with retroviruses during the course of evolution. Most seem to be harmless — over time they have accumulated mutations that destroy their ability to give rise to replicating virus, and they occur at the same location in the genome of all individuals of a species. But others are more recent; they show genetic polymorphism and potentially encode infectious virus.

In general, endogenous proviruses are not highly expressed, although occasional transcription can occur. Due to adaptations of host viral-receptor proteins, many otherwise intact endogenous retroviruses cannot infect cells derived from their current host. However, viral replication can take place in cells from a different species; that is, the viral host range is 'xenotropic'. So, if grown in apparently retrovirus-free mice, human cells will often show signs of infection by a mouse xenotropic virus⁵. A pig-to-human xenograft is conceptually similar, prompting fears that a porcine retrovirus from a transplanted organ might infect human recipients, with unknown consequences⁶.

Until recently, porcine retroviruses received little attention: retrovirus production was seen from a number of pig cell lines, but these viruses seemed to be ecotropic in host range — they infected only pig cells⁷. Patience *et al.*⁴ reinvestigated the infectivity of virus that was produced by two cell lines

derived from pigs, MPK and PK-15. They confirmed that MPK cells make ecotropic virus; however, the PK-15 cells seemed to be making a mixture of ecotropic and xenotropic viruses. Moreover, the xenotropic virus could replicate, albeit very inefficiently, in a variety of human-derived cell lines (although primary cultures of human cells have not yet been examined in any detail).

Limited sequence-studies of the viral polymerase gene indicate that the ecotropic and xenotropic viruses are very similar to one another, and that they belong to the mammalian C-type family of retroviruses. Hybridization studies predict that the pig genome contains around 50 related proviruses, one or more of which is expressed in a variety of tissue types, including heart and kidney — the two organs most likely to be transplanted. Unfortunately no data are presented concerning the infectivity or host range of the spontaneously expressed virus. Also absent from the paper are sequence data from the region of the virus which determines host range, and information about the number of endogenous xenotropic proviruses in the pig genome. Indeed, although unlikely, it is still possible that none of the endogenous proviruses is xenotropic, and that the xenotropic PK-15 virus contains sequences derived from a non-porcine retrovirus.

Evaluating the risk posed by viruses of the sort described by Patience *et al.* is not simple. The first question is whether transmission can occur. It will be interesting to see whether studies using the polymerase chain reaction, with primers designed to the pig retrovirus, show any evidence for transmission in primates transplanted with pig organs, or in humans with occupational exposure to pigs or pig tissues. Frequent contact between pigs and humans has apparently never led to retrovirally induced disease, possibly because human complement can lyse retrovirus that is grown in animal cells⁸. But under the circumstances of a transplant — close proximity and long-term contact — this barrier to spread could probably be overcome, especially as the measures taken to prevent hyperacute rejection and permit xenotransplantation will reduce the efficiency of lysis. Patience *et al.* also show that porcine retrovirus grown in human cells is no longer susceptible to lysis by human complement, implying that once viral breakout occurs there will be no further barrier to spread by this resistance mechanism.

If a viraemic state is established, a number of outcomes are possible. Most endogenous xenotropic retroviruses are essentially non-pathogenic⁹, so transmission of virus would

probably have no consequences for the organ recipient or for society. But oncogenicity trials of xenotropic virus in immunosuppressed primates need be considered in light of a report that a retrovirus resulting from recombination between two murine retroviruses can cause lymphomas in macaques¹⁰. Even if a transmitted virus gives rise to tumours at a low frequency, risk-benefit analysis might suggest that this is a chance worth taking, as long as the virus is confined to the transplant recipient. Although less likely, it is much more frightening to consider that the initial virus could spread to the general population, possibly following mutational or recombinational events.

In the long run, the best way to rule out the threat posed by the transmission of porcine retroviruses would be to ensure that herds of donor pigs did not carry proviruses which could infect human cells. A genetic approach should be feasible, using the same general strategy as described for studies of murine leukaemia viruses^{11,12}. By comparing the host-range-determining sequences of ecotropic and xenotropic proviruses, it should be possible to prepare probes specific for each viral subgroup. These probes would allow the inheritance and distribution of xenotropic proviruses to be followed in pigs, as well as tests

to show whether the viruses are expressed in transplanted organs. The polymorphism of endogenous proviruses in mice¹² gives us good reason to believe that conventional breeding techniques could then be used to generate pigs that are free of those endogenous proviruses that give rise to concern. Although this represents a challenging and time-consuming task, under the circumstances there seems to be no alternative. Even so, the lurking fear of novel, undiscovered retroviral flies in the oinkment would persist. □

Jonathan P. Stoye is at the National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

1. Cozzi, E. & White, D. J. G. *Nature Med.* **1**, 964-966 (1995).
2. Sharma, A. et al. *Proc. Natl Acad. Sci. USA* **93**, 7190-7195 (1996).
3. Kennedy Committee *Animal Tissues into Humans: A Report by the Advisory Group on the Ethics of Transplantation* (1997).
4. Patience, C., Takeuchi, Y. & Weiss, R. A. *Nature Med.* **3**, 282-286 (1997).
5. Tralka, T. S. et al. *J. Natl Cancer Inst.* **71**, 591-599 (1983).
6. Stoye, J. P. & Coffin, J. M. *Nature Med.* **1**, 1100 (1995).
7. Lieber, M. M., Sherr, C. J., Benveniste, R. F. & Todaro, G. J. *Virology* **66**, 616-619 (1975).
8. Takeuchi, Y. et al. *Nature* **379**, 85-88 (1996).
9. Levy, J. A. *Curr. Top. Microbiol. Immunol.* **79**, 111-213 (1978).
10. Vanin, E. F., Kaloss, M., Broscius, C. M. & Nienhuis, A. W. J. *Virology* **68**, 4241-4250 (1994).
11. Stoye, J. P. & Coffin, J. M. *J. Virol.* **61**, 2659-2669 (1987).
12. Frankel, W. N., Stoye, J. P., Taylor, B. A. & Coffin, J. M. *Genetics* **124**, 221-236 (1990).

Crystallography

The negative side of crystal growth

Paul Calvert and Stephen Mann

What controls the growth of inorganic minerals, such as bone, in the body? A paper by Yamashita et al.¹ shows that the calcium phosphate mineral apatite grows rapidly on the negative-pole but not on the positive-pole surface of ferroelectric crystals. It may be that uniform electric fields, rather than the localized charges usually cited, determine the sites of crystal nucleation and overgrowth.

There are two key questions in biomineralization: what characteristics of organic and inorganic surfaces promote the nucleation of crystalline overgrowths of calcium carbonate or phosphate from supersaturated solutions; and what causes this inorganic layer sometimes to adopt a specific crystallographic orientation with respect to the organic surface? The popular answer to both questions is that this is an epitaxial effect; that is, the spacings between atoms in the substrate induce the growth of the mineral layer in a specific orientation for which the atom spacings almost match across the interface. But it has been evident for some time that this concept does not do a very good job of explaining what actually works, and so stereochemical and charge matching have been suggested as contributing factors^{2,3}.

Yamashita and colleagues have found

instead that what matters is the bulk electrical polarization of substrates such as calcium titanate, barium titanate and hydroxyapatite (which occurs in bone). This observation is important because there is considerable interest in inducing apatite mineralization on synthetic surfaces in order to promote bone attachment to orthopaedic implants. And there have also been efforts to connect the known piezoelectricity of bone with electrical effects on bone healing and with bone remodelling under stress. More prosaically, there are a host of practical problems associated with inorganic mineralization in boilers and circulating water systems that might benefit from the new results.

Ferroelectric crystals, named in analogy to ferromagnetism, contain aligned electric dipoles arranged in randomly oriented domains, so that a normal crystal shows no overall polarity. In a strong electric field, typically a few hundred volts per millimetre, the domains realign parallel to the field and the crystal develops a bulk polarity. Squeezing or hitting such a crystal produces a surface charge; this piezoelectric effect is used in gas barbecue lighters and microphones⁴.

A polarized ferroelectric crystal has a strong field at the surface owing to the aligned dipoles throughout its structure, and

this should attract a layer of ions from a surrounding solution. However, each individual dipole is not large, typically corresponding to a charge displacement of 0.1 Å, or just 1/30th of an atom-atom spacing⁵. So the crystal surface will present a broad uniform field of attraction. This is very different from an electrostatically charged crystal surface, which has a periodic array of localized binding sites for ions of opposite polarity.

Yamashita and colleagues immersed ferroelectric crystals for periods of between several hours and a week in '1.5 × simulated body fluid' — a saline solution of phosphate, sulphate, bicarbonate, magnesium and calcium made to 1.5 times the concentration in the body^{6,7}, and heated to 300 °C. The rate of crystal overgrowth on apatite polarized by a field of 120 V mm⁻¹ was six times that observed on a non-polarized substrate. Furthermore, the accelerated crystal growth was only observed on the negative-pole face. These observations are reminiscent of the claims for control of mineralization (and, less plausibly, many other things) by magnets in water systems.

As an explanation, the authors postulate that the negative crystal surface attracts calcium ions, increasing the local nucleation rate of apatite from solution. One would also expect the negative surface to deplete phosphate and the positive surface to concentrate phosphate, reversing the effect, unless the relatively low charge density of the phosphate anions means that they cannot compete effectively with chloride adsorption.

Another possibility is that Ca²⁺ ions clus-

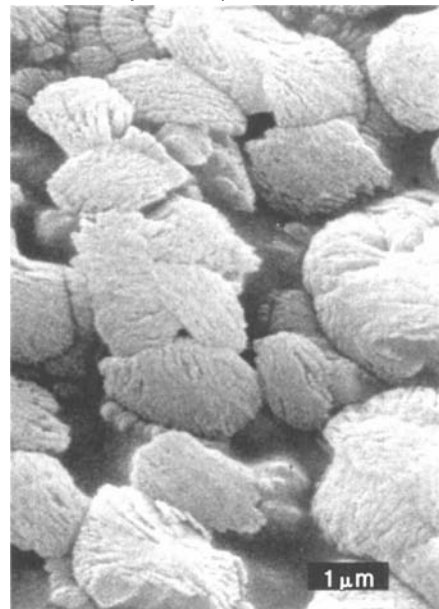


Figure 1 Calcium oxalate on a thin film of acrylic copolymer. As a test of the idea that local charges control nucleation, a series of films of acrylic copolymers were tested for mineralization from supersaturated calcium oxalate solutions¹¹. Neutral, hydroxylated polymers were unexpectedly found to be much more effective than charged, carboxylated films.

ter on the negative surface of the ferroelectric crystal, leading to charge reversal and attracting phosphate anions from solution. That would increase the lifetime of incipient aggregates and therefore lower the activation energy of nucleation. But why doesn't this work with an ordinary surface charge? Perhaps such a periodic spatial charge pins the Ca^{2+} ions at lattice sites of the substrate, preventing the incipient clusters from relaxing into the crystal structure of calcium phosphate; in contrast, the gentler polarization field of the ferroelectric crystal may allow the clusters the extra freedom required.

The overgrowths on the negative face of these ferroelectric substrates have no preferred orientation, but they may do in other systems. A uniform electric field could influence the pathway followed by inorganic clusters during structural relaxation to the overgrowth phase. This idea could be tested by a combination of theoretical modelling and further experimental studies — possibly with calcium carbonates, because that system is intrinsically simpler and good computer codes are available.

A problem with the lattice-match model has always been that the interatomic spacings are not variable. Hence it is not really possible to test this model by changing spacing without also changing surface chemistry. The ferroelectric effect, on the other hand, seems to provide abundant possibilities for experimental tests of any theory, simply by changing crystal type and microstructure or the external field.

It will be interesting to discover whether more subtle dipole-induced fields can influence inorganic crystallization. The dipole moment generated by an α -helical polypeptide has been proposed as a possible explanation for the antifreeze activity of certain proteins from polar fishes⁸; crystals of ice grown under compressed monolayers of long-chain alcohol surfactants are oriented according to networks of interfacial hydrogen bonds⁹; similar fields are probably responsible for the oriented nucleation of the hydrated face of gypsum in the presence of the same monolayers¹⁰; and during bone formation, the clustering of ions and water in the intramolecular spaces of aligned collagen molecules might also be influenced by non-periodic fields.

Although non-periodic fields are weak compared with electrostatic fields, even small changes in the local interfacial energy could be enough to markedly influence the local energy states of incipient nuclei and their transformation into mineral phases. Hetero-epitaxy is subtle and mysterious. A new focus on fields may solve what we have failed to explain by concentrating on spacings. □

Paul Calvert is in the Department of Materials Science and Engineering, University of Arizona, Tucson, Arizona 85721, USA. Stephen Mann is in the School of Chemistry, University of Bath, Bath BA2 7AY, UK.

1. Yamashita, K., Oiawa, N. & Umegaki, T. *Chem. Mater.* **8**, 2697–2700 (1996).

2. Addadi, L., Moradian, J., Shay, E., Maroudas, N. G. & Weiner, S. *Proc. Natl Acad. Sci. USA* **84**, 2732–2736 (1987).
3. Mann, S. *Nature* **332**, 119–124 (1988).
4. Jona, F. & Shirane, G. *Ferroelectric Crystals* (Dover, New York, 1993).
5. Kittel, C. *Introduction to Solid State Physics* (Wiley, New York, 1968).
6. Kokubo, T. *Eur. J. Solid State Inorg. Chem.* **32**, 819–827 (1995).

7. Abe, Y., Kokubo, T. & Yamamuro, T. *J. Mater. Sci.: Mater. Med.* **1**, 233–238 (1990).
8. Yang, D. C. S., Sax, M., Chakraborty, A. & Hew, C. L. *Nature* **333**, 232–237 (1988).
9. Majewski, J. *et al. Science* **261**, 899–902 (1993).
10. Douglas, T. & Mann, S. *Mater. Sci. Eng. C* **1**, 193–199 (1994).
11. Burdon, J. *et al. Mater. Sci. Eng. C* **4**, 133–137 (1996).

Biomembranes

Lipids beyond the bilayer

Ben de Kruijff

Cells and organelles are enclosed by membranes, which not only define their contents but also act as a filter and support active transport between the external and internal environments. Such functions are carried out by specific proteins embedded within a lipid bilayer, and understanding the way that these proteins operate requires insight into their three-dimensional structure. Despite some pioneering studies using two- and three-dimensional electron and X-ray crystallography, however, crystallization of membrane proteins has remained largely a dream (and, for some, pursuing that dream has been a nightmare).

As described in *Proceedings of the National Academy of Sciences*, Landau and Rosenbusch¹ have devised a novel approach to the problem using lipidic cubic phases; these are ordered, three-dimensional water–lipid systems, in which the lipids are organized in a highly curved, interwoven bilayer network (Fig. 1). The significance of the study goes beyond opening a new avenue towards the elucidation of the structure of membrane proteins. It also provides a clue to the long-standing question as to the functional importance of non-bilayer lipids in membranes.

So far, crystals suitable for structural analysis have either been produced by three-dimensional crystallization of membrane proteins from detergent micelle solutions or by two-dimensional crystallization within planar lipid bilayers. Each of the few success stories to date — crystallization of a bacterial photoreaction centre², and of plant³ and bacterial⁴ light-harvesting complexes — has required a specific set of crystallization conditions. The approach followed by Landau and Rosenbusch is entirely different. They created a membrane-mimicking system consisting of a lipidic cubic phase containing the membrane protein bacteriorhodopsin. This system allowed the formation of three-dimensional protein crystals which were amenable to X-ray crystallography. The rationale offered by the authors is that the large lipid–water interface of the cubic phase provides nucleation sites, while the network of bilayers allows the protein molecules to diffuse laterally to the crystallites and let them grow. The resulting crystals diffracted to 3.7 Å resolution.

This approach can also be viewed from

another angle, one that sheds light on the puzzle of polymorphism among membrane lipids. The first pieces of this puzzle were identified some 30 years ago by several research groups, notably Luzzati and collaborators⁵, who found that lipids generally can self-assemble into different aggregate structures including bilayer but also non-bilayer structures. In the 1970s and early 1980s, it was discovered that such polymorphism also applies to membrane lipids — biomembranes contain a substantial amount of lipids which, on their own, prefer to organize themselves in non-bilayer structures⁶. Moreover, it turned out that these non-bilayer lipids are precisely regulated, implying that they are of considerable functional importance^{7,8}.

What has remained unclear is why biomembranes should contain so much non-bilayer lipid. The overall organization of the membrane is as a bilayer, and it would seem that only a few lipid molecules should be

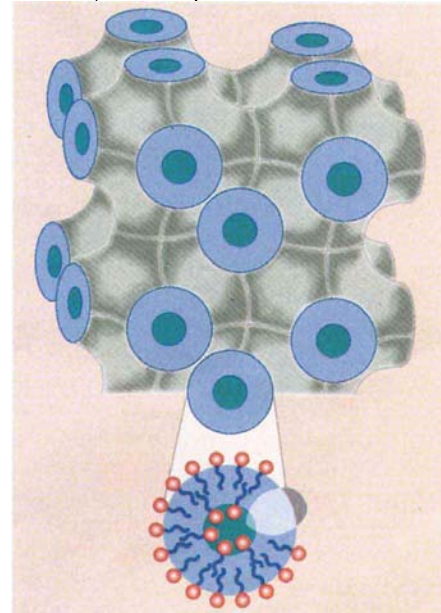


Figure 1 Schematic representation of a lipidic cubic phase. That produced by Landau and Rosenbusch¹ consisted of monoolein (lipid), water and bacteriorhodopsin (membrane protein). The membrane system is interpenetrated by water channels, shown here in darker blue. Detail of the curved lipid bilayer around a water channel, including a membrane protein molecule, is shown below. (Redrawn from ref. 1 with acknowledgement to ref. 13.)

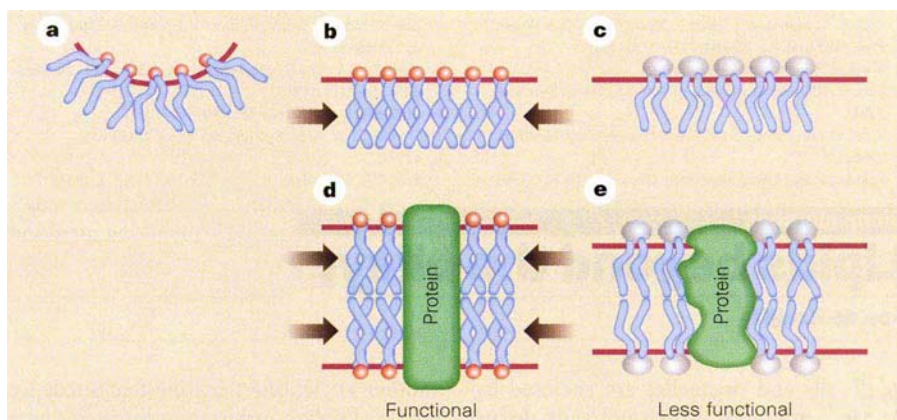


Figure 2 A model for the way in which non-bilayer lipids (red headgroups) preserve the functional structure of membrane proteins (green), as described in the text: this view, based on the results of Landau and Rosenbusch¹, helps to explain why some biomembranes contain surprisingly large amounts of non-bilayer lipids. As shown in d, non-bilayer lipids exert high lateral packing pressure (arrows) on membrane proteins, keeping them (literally) in better shape than do bilayer-preferring lipids (e; grey headgroups).

necessary to build a (transient) non-bilayer structure for, for instance, the formation of such specialized membrane structures as the tight junction⁹. But take, for example, the membrane of the bacterium *Escherichia coli*, in which the lipids are predominantly (and perhaps completely) in a bilayer structure; yet 70 per cent of the phospholipids in the membrane are in the form of phosphatidylethanolamine (PE), which has a strong tendency to organize in non-bilayer structures.

A step forward in tackling these issues came with the construction, by Dowhan and colleagues, of *E. coli* mutants lacking PE. The results were surprising. In the mutants, membrane integration and functioning of important transport proteins such as lac permease were severely impaired¹⁰, and protein translocation was much reduced¹¹; reintroduction of non-bilayer lipids resulted in the specific restoration of translocation. So these studies further implicated non-bilayer lipids in maintaining the functional structure of integral membrane proteins such as transporters.

So how does Landau and Rosenbusch's study fit into this story? A straightforward explanation presents itself. The cubic phases they used are structures that are intermediates between a bilayer and non-bilayer organization, and can only be formed with substantial amounts of non-bilayer lipids. It is therefore to be expected that the functional structure of a protein will be better preserved within the cubic phase; and because native proteins crystallize best, the crystals grow.

Taking the view that non-bilayer lipids somehow stabilize membrane proteins helps to rationalize some long-standing observations — for instance, that non-bilayer lipids are often required for functional reconstitution of membrane proteins; and that the more functional properties a membrane has, the more non-bilayer lipids it contains. A possible molecular explanation for the general phenomenon is shown in Fig. 2. The non-bilayer lipids have a strong tendency to or-

ganize in aggregate structures with a concave interface because of their relatively small headgroups (a). Within the constraints of a bilayer or biomembrane (b), this results in a high lateral packing pressure in the hydrophobic core of the membrane (arrows). Bilayer-preferring lipids do not give rise to such forces because their headgroup size is more equivalent to the cross-sectional area of the hydrocarbon region of the lipid (c). The internal lateral packing force presses against the outer surface of membrane proteins and thereby keeps them in a functional state (d). In the absence of such force, the protein relaxes and becomes less efficient or non-functional (e).

Landau and Rosenbusch's results will both encourage crystallographers interested in membrane proteins and guide those working on lipid polymorphism in designing new experiments. The striking observation that PE is a well-defined component of the crystals of the membrane protein cytochrome *c* oxidase¹² adds to the expectation that we will learn a lot more about non-bilayer lipids over the next few years. □

Ben de Kruijff is in the Centre for Biomembranes and Lipid Enzymology, Department of Biochemistry of Membranes, Utrecht University, PO Box 80.054, 3508 TB Utrecht, The Netherlands.

1. Landau, E. M. & Rosenbusch, J. P. *Proc. Natl. Acad. Sci. USA* **93**, 14532–14535 (1996).
2. Deisenhofer, J. et al. *Nature* **318**, 618–623 (1985).
3. Kühlbrandt, W., Wang, D. N. & Fujiyoshi, Y. *Nature* **367**, 614–621 (1994).
4. McDermott, G. et al. *Nature* **374**, 517–520 (1995).
5. Luzzati, V., Gulik-Krzywicki, T. & Tardieu, A. *Nature* **218**, 1031–1034 (1968).
6. Cullis, P. R. & de Kruijff, B. *Biochim. Biophys. Acta* **559**, 399–420 (1979).
7. Wieslander, A., Christiansson, A., Rålfors, L. & Lindblom, G. *Biochemistry* **19**, 3650–3655 (1980).
8. Goldfine, H. & Johnston, N. C. *Biochemistry* **26**, 2814–2822 (1987).
9. Kachar, B. & Reese, T. S. *Nature* **296**, 464–466 (1982).
10. Bogdanov, M., Sun, J., Kaback, H. R. & Dowhan, W. *J. Biol. Chem.* **271**, 11615–11618 (1996).
11. Rietveld, A. G., Koorengevel, M. C. & de Kruijff, B. *EMBO J.* **14**, 5506–5513 (1995).
12. Tsukihara, T. et al. *Science* **272**, 1136–1144 (1996).
13. Lindblom, G. & Rålfors, L. *Biochim. Biophys. Acta* **988**, 221–256 (1989).

Daedalus

Think the good thought!

Last week Daedalus proposed an ion cyclotron resonance encephalograph. The ions exchanged by a nerve impulse are deflected into orbits by a magnetic field; where their orbital frequency matches that of an applied a.c. electric field, they absorb that frequency in resonance. The electric and magnetic fields have gradients such that the resonance conditions are met only in a small chosen volume of the brain. The new machine resolves tiny centres of activity, and times them to milliseconds.

With its aid, Daedalus is now exploring the brains of DREADCO volunteers. He is locating the origins of sensations, actions, thoughts and feelings of all sorts, and identifying their resonant signatures. He should then be able to detect, reward and punish thoughts themselves, by direct pavlovian conditioning. An ionic encephalograph feeding a computer could scan the brain of the subject for secret racist or sexist thoughts, sceptical views on democracy or ecology, or resentment at the government's rightful constant demands for money and information; and could punish them instantly by, for example, an electric shock to the ears. The subject would rapidly learn not to have such anti-social ideas. This pavlovian process would be the first effective punishment in the history of jurisprudence. The swelling ranks of deviants and subversives could at last be mentally conditioned into responsible, compliant citizens.

The next step, of course, would be not merely to detect the thoughts already in the subject's brain, but to inject approved thoughts into it. Sadly, this seems unfeasible. Even very intense fields could not force the nerves in the resonant region to fire, nor to fire in the right complex patterns. They might, however, be slightly sensitized. This mental 'tickling' would at least encourage the right answer or the proper response, which once elicited could be rewarded. So education would take a giant leap forward, too.

Even better, the system could be used in biofeedback. By showing the subject what his brain was doing right or wrong, he could learn for himself far faster than any teacher could teach him. Thus, in learning proper speech, the encephalograph could be targeted on Broca's area of the brain, and programmed to identify each word as the subject chose it, and drive an audio memory to speak it. However, the brain operates hundreds of milliseconds ahead of the conscious mind. The subject might have the unnerving experience of hearing the machine saying his words ahead of him.

David Jones

The use and abuse of climate models

Kevin E. Trenberth

Projections of future climate change depend largely on the results of computer models. Such models are becoming increasingly sophisticated, but they do not offer the certainties that policy-makers would like.

Humankind is performing a great geophysical experiment¹. By modifying the Earth's environment in various ways, we are changing the climate. The extent and the rate of these changes are unclear, as is what (if anything) should be done about them, but that the experiment is underway is not in doubt. The environmental changes of most relevance are in land use (farming, building cities), storage and use of water (dams, reservoirs, irrigation), generation of heat, and — most notably — the burning of fossil fuels.

In particular, fossil-fuel combustion pollutes the atmosphere and alters the balance of radiation on Earth through both visible particulate pollution (called aerosols), and gases that change the composition of the atmosphere. These are known as greenhouse gases because they are relatively transparent to incoming solar radiation, but absorb and re-emit outgoing infrared radiation, thus creating a blanketing effect which results in warming. For example, as a consequence of human activities, carbon dioxide concentrations in the global atmosphere^{2–4} have increased by about 30 per cent over pre-industrial values. Global warming and associated climate change is expected as a result, and the global mean temperatures have indeed risen over the past hundred years⁵ (Fig. 1).

If this experiment turns out badly — however that is defined — we cannot undo it. We cannot even abruptly turn it off, because too many of the things we are doing now have long-term ramifications. For instance, carbon dioxide has an atmospheric lifetime of over a century⁵ and simply stopping increases in emissions would still result in increases in atmospheric concentrations for many decades. The only way to reverse those trends is to reduce emissions to well below current levels^{4,5}. Moreover, changes underway in the oceans would endure, because of the oceans' huge heat capacity.

If we had two planet Earths, identical in every respect except that the residents of one adopted measures to avoid polluting the atmosphere while residents of the other did not, we could see how the climates of the two planets would diverge, and what the consequences would be. But we don't and we

can't. Instead we have to do the next best thing — try to understand the climate system well enough to build a good model of the planet Earth system and use this model to perform the experiments. We can indeed construct a miniaturized physical model of the Earth–Sun system; but we cannot readily include the effects of gravity, and the rich complexity of the atmosphere and oceans. The alternative is to build a virtual model of the Earth in a computer.

The models

These computer models are based upon physical laws represented by mathematical equations and expressions that are solved using numerical methods as applied to a three-dimensional grid over the globe. The most complete versions are referred to as 'Earth system models' and those which deal exclusively with climate are called 'climate models' or, if they are very comprehensive, 'climate system models'⁶.

But how useful are these models in making projections of future climate? Opinion is polarized. At one extreme are those who take the model results as gospel; at the other are those who denigrate such results simply because they distrust models, or on the grounds that the model performance is obviously wrong in some respect or that a process is not adequately included. The truth lies in between. All models are of course wrong because, by design, they depict a simplified view of the system being modelled. Nevertheless, many — but not all — models are very useful.

A full climate system model⁶ should deal with all of the physical, chemical and biological processes that occur in nature and the

interactions among the components of the climate system (Fig. 2, overleaf). A major component of them is the so-called atmospheric general circulation models (AGCMs); these are designed to simulate the detailed evolution of weather systems and weather phenomena, as well as the physical and dynamical processes involved. For AGCMs, typical resolutions for climate simulations are about 250 km in the horizontal direction and 1 km in the vertical. These models are widely used and tested every day in making weather forecasts, although with finer resolution, and have predictive value out to about ten days ahead⁷.

The theoretical limit to the predictability of weather is about two weeks, which stems from the phenomenon of chaos — small uncertainties in the analysis of current weather conditions rapidly grow and eventually become large enough to make the forecast worthless⁸. This essentially random-error component is overcome in climate forecasts by predicting only the statistics of the weather (that is, the climate). Accordingly, systematic influences of changing conditions in the ocean, sea ice, land surface, solar radiation or other factors are reflected in atmospheric variations on various time scales. The most obvious example is the climate change with the seasons. An ensemble of several climatic simulations, each of which begins from somewhat different starting conditions, can be used to establish which climatic features are reproducible in the simulations and thus are predictable with the model.

These features constitute the climate signal, while those which are not reproducible can be considered weather-related climate noise. Climate predictability is a function of spatial scales. Natural atmospheric variability is enormous on small scales and most effects on climate (forcings), such as from increases in greenhouse gases, are predominantly global. So the noise level of natural variability will mask a climate signal more as smaller regions are considered⁵.

Model errors

The latest coupled atmosphere–ocean–land–sea-ice models provide very good simulations of average climate conditions and

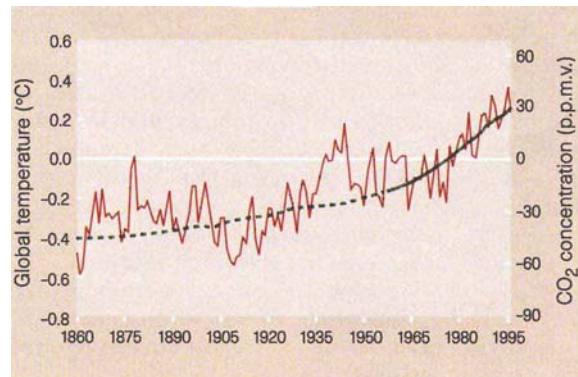


Figure 1 Estimated changes in annual global mean temperatures⁵ (red) and carbon dioxide (green) over the past 137 years relative to a 1961–90 base period. Earlier values for carbon dioxide are from ice cores³ (dashed line), and for 1957 to 1995 from direct measurements made at Mauna Loa, Hawaii². The scale for carbon dioxide is in parts per million by volume (p.p.m.v.) relative to a mean of 333.7 p.p.m.v.

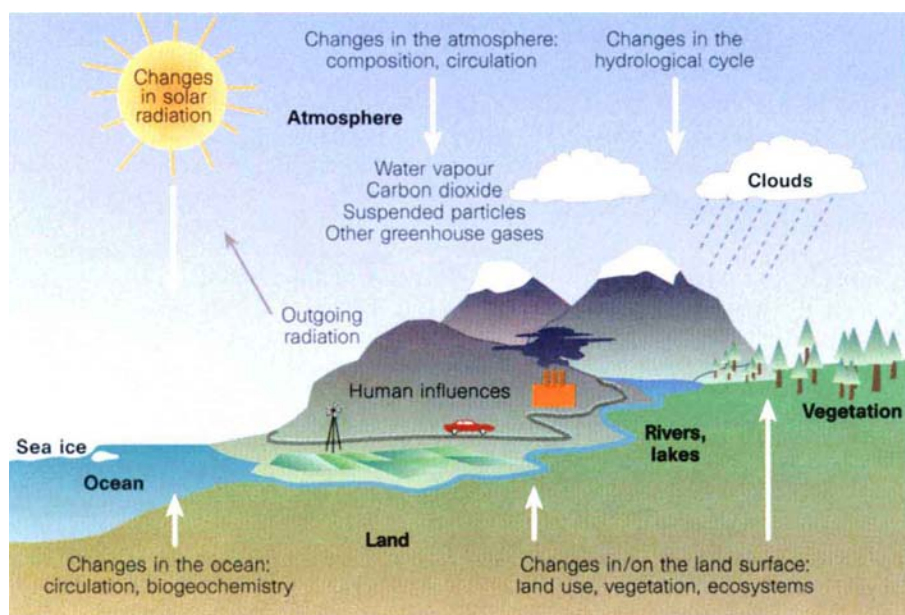


Figure 2 Components of the global climate system — atmosphere, ocean, sea ice, land surface, surface hydrology and biosphere — their processes and interactions, and some aspects that may change.

their evolution with the seasons. Nevertheless, climate models do contain obvious errors in some features when compared with observations. The climate modeller is then faced with a decision. The 'error fields' contain useful information about the performance of the model. Experiments performed with such a model have the advantage that they consistently include all the physical processes, but the disadvantage is that the systematic errors may distort the results because many feedback processes change in strength as the temperatures change. For example, if a model atmosphere is simulated to be too cold by 4 °C (not uncommon a few years ago)⁹, the water-holding capacity of the atmosphere is typically reduced by about 20 per cent thereby greatly influencing evaporation and precipitation; places that should be receiving rain will instead have snow.

An alternative strategy is to include an artificial fix known as 'flux adjustment' to keep the simulated climate closer to that observed. The exchanges (fluxes) of energy, water and momentum between the model ocean and atmosphere are adjusted so that the modelled sea surface temperatures and other surface fields are close to those observed. These fixes are then held constant in any experiments. There are merits in approaches both with and without flux adjustment and the results can be compared, but both sets of results are open to the criticism that the model is clearly deficient in some ways. So a lot of research is focused on developing climate models that greatly reduce systematic errors and eliminate the need for flux adjustment.

To further reduce the effects of any model errors on results, a strategy has been designed for carrying out climate experi-

ments which removes much of the effects of these errors and flux adjustments. First, a 'control' climate simulation is run with the model. Then the climate-change simulation is run, for example with increased carbon dioxide in the model atmosphere. Finally the difference is taken to provide an estimate of the change in climate. This differencing technique removes the effects of flux adjustment, as well as the systematic errors that are common to both runs. But comparison of different model results makes it apparent that the nature of some errors influences the outcome, so that complicated feedback effects do take place.

An example of a problem that cannot be alleviated by this approach occurs, for instance, if the control climate produces no rainfall in a monsoon area where it should occur (the monsoon rains may be in the wrong location). Then it is impossible for the rainfall to be reduced in a climate experiment and the only possible outcome is an increase in rainfall.

Although it is desirable for a model to be as realistic as possible, this is not always feasible. Indeed, a full model of the climate system would be just as complex as the system itself and almost as difficult to understand, except that complete model datasets could be created for analysis and experiments could be performed. Some processes or influences are so complex and so poorly understood, or simply cannot be resolved by the scales represented in a model (which is related to computer limitations), that it may be better to leave them out altogether. In other cases, attempts are made to include the average influences using a physically based 'parametrization', in which unresolved processes are represented through resolved variables.

Probably the single greatest uncertainty in climate models stems from their treatment of clouds¹⁰. The enormous variety of cloud types, their variability on all space scales (ranging from sub-millimetre to thousands of kilometres) and time scales (microseconds to weeks) poses a special challenge, particularly in depicting their influence on incoming solar and outgoing infrared radiation and their role in precipitation.

Model validation

Climate models that have been developed thus far for application to the greenhouse-gas problem have largely centred on the physical climate system. Typically, the concentrations of constituents of the atmosphere, including radiatively important species such as ozone and carbon dioxide, have either been fixed or specified as varying functions of time. In such 'scenarios', the concentrations of the gases do not depend on the climate changes going on in the model even though, in nature, changes in rainfall or temperatures may profoundly affect the sources and sinks of some of the greenhouse gases. Similarly, land surface processes have been greatly oversimplified in the models, and biological, ecological and chemical processes may not be included at all. Nevertheless, these approximations and omissions are appropriate for addressing certain scientific questions.

In all circumstances 'sensitivity tests' should be carried out to check how sensitive the result is to small changes in what is done. For instance, simplification of land surface processes seems to be justified for very large spatial scales, but not for studying regional effects. To explore all possible scenarios and the effects of approximations and assumptions, simpler models are also widely used. They are 'tuned' to the more complex climate models and have the advantage of using much less computer time. All of these models may be useful tools provided their limitations are properly taken into account.

Model results should be judged by considering all the assumptions (such as certain things being held constant) and approximations involved. It is generally inappropriate to take the model result at face value because it must be comprehensively evaluated. Thus the model performance in simulating the annual cycle, interannual variability, the past climate record including what is known about climates of the distant past (palaeoclimates), and in simulating responses to a volcanic eruption, must also be factored into how much weight is given to the result.

This process requires comparison of the model to observations and the assembling of the necessary data sets. It also requires initializing the model with observations of

the current or starting state of many model variables, especially those which vary over many years (land surface, sea ice, ocean). If flux adjustment is applied to a model then some results may be deceptively good and the magnitude of the flux adjustments must be factored into the evaluation. If the simulations are judged to be realistic, the model may prove to be a useful tool for applications to scientific, economic and social questions, including making assessments and projections of possible future climates.

The best models encapsulate the current understanding of the physical processes involved in the climate system, the interactions between them, and the performance of the system as a whole. They have been extensively tested and evaluated using observations. The complexity and nonlinearity of the climate system is such that good models often provide the only means of quantifying the result of a perturbation to that system.

As regards alternative approaches, statistical and empirical results are not reliable because the possible outcomes usually fall well outside the realm of conditions previously experienced, and the observed record on which to base such extrapolations is far too short; moreover, projections about how the climate may or may not change based upon indirect experience or intuition are effectively making use of a vastly cruder model. It is easy to criticize a model. But the challenge to any proponent of, say, a missing process is to insert it into one of these models (several are available) and demonstrate that it makes a quantitative difference to the outcome. The burden of proof that a

model result is not valid should be on the critic, not the modeller.

Future developments

Several major laboratories around the world are hard at work on model development, and impressive progress is being made. Necessarily, such research is multidisciplinary and involves experts in many different areas collaborating to produce a climate system model that is more than just the compilation of the components. Because it takes thousands of years for a model that includes a deep ocean to come into equilibrium, integration of a full climate model can use all of the computer time available on the biggest and fastest computers. For example, the climate system model at the National Center for Atmospheric Research¹¹ (NCAR, the institution for which I work) runs with an AGCM resolution of 2.8° latitude/longitude and 18 levels in the vertical; it has a non-eddy resolving ocean model with 45 levels, and a coarse horizontal resolution (nominally 2°); and it includes dynamical sea ice and land surface components. It takes about 3.5 hours on a Cray C90 machine with 16 dedicated processors to run for one simulated year.

Computing power is one key to future progress. To carry out comprehensive numerical climate experiments and understand the results, a supercomputer has to be dedicated to the task for many months. That is an ideal that is seldom achieved. The NCAR model has been run for over 200 simulation years and has a stable climate quite close to the observed without flux adjustment. This has been achieved through improvement in representations of physical processes in both the model atmosphere and ocean, which is the second key to progress.

Projections of global warming

But what of the issue that most exercises the public — the changes in atmospheric composition arising from human activities and the consequent effects on climate? The Intergovernmental Panel on Climate Change (IPCC; ref. 5) has made projections of future global warming based upon model results to the year 2100. Because human actions are not predictable in any deterministic sense, such projections necessarily contain a 'what if' element to account for differing patterns and amounts of pollution emission in the future. Results depend greatly upon these scenarios which detail the expected rates of increases of the greenhouse gases and aerosols.

In addition, for a given scenario, the rate of increase in global mean temperature depends on the model and features such as how clouds are depicted, so that for a doubling of carbon dioxide concentrations, for example, typical equilibrium global warming ranges from 1.5 to 4.5 °C with a best esti-

mate of 2.5 °C. Natural variability also blurs the picture from a single model result, producing a scatter of about ± 0.2 °C. Accordingly, the projections for a mid-range emissions scenario, in which carbon dioxide concentrations increase from 350 p.p.m.v. (parts per million by volume) in 1990 to 700 p.p.m.v. in 2100 produce global mean temperature increases ranging from 1.3 °C to 2.9 °C above 1990 values with a best estimate of about 2 °C (Fig. 3; ref. 12). Although these projections include crude estimates of the effects of sulphate aerosol, they deliberately omit other possible human influences such as changes in land use. Moreover, many other climate changes accompany global warming, such as altered extremes of rainfall and the availability of fresh water, and these may have much greater consequences for human society. A major concern is that the rates of climate change as projected exceed anything seen in nature in the past 10,000 years.

Statements such as these, given with appropriate caveats, are likely to be the best that can be made because they factor in the substantial understanding of many processes included in climate models. Such projections cannot offer certainty, but they are far better than declaring ignorance and saying nothing at all. If they prove to indicate sufficiently adverse effects on climate, then it is generally assumed that action will be taken to prevent those effects. What constitutes 'sufficiently adverse' is of course a matter for debate, as is what, if any, remedial measures should be taken. Such debates are not just for scientists, but for politicians and the peoples of the world whom they serve — and they have to be based on an appreciation of both the strengths and the weaknesses of climate models. □

Kevin E. Trenberth is in the Climate Analysis Section, National Center for Atmospheric Research (NCAR), PO Box 3000, Boulder, Colorado, USA (e-mail: trenberth@ncar.ucar.edu). NCAR is sponsored by the National Science Foundation.

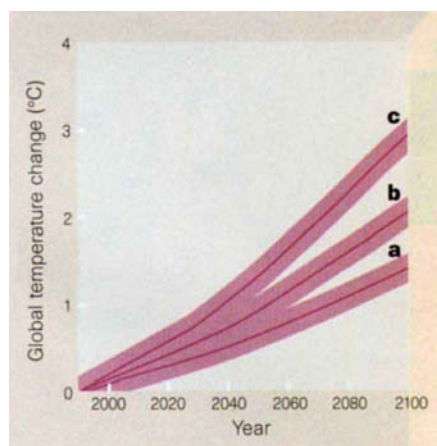


Figure 3 Projected changes in global mean temperature from 1990 to 2100, corresponding to an emissions estimate¹² involving increases in both carbon dioxide from 350 to 700 p.p.m.v. and in sulphate aerosol (a 'mid-range scenario'). The three curves show the average changes and the scatter (shading) from natural variability corresponding to models with low (a), best estimate (b) and high sensitivity to change (c).

1. Revelle, R. & Suess, H. E. *Tellus* 9, 18–27 (1957).
2. Keeling, C. D. et al. in *Aspects of Climate Variability in the Pacific and Western Americas* (ed. Peterson, D. H.) 165–236 (Geophys. Monogr. 55, Am. Geophys. Un., Washington DC, 1989).
3. Raynaud, D. et al. *Science* 259, 926–934 (1993).
4. Schimel, D. et al. in *Climate Change 1994* (eds Houghton, J. T. et al.) 35–71 (Intergovernmental Panel on Climate Change/Cambridge Univ. Press, 1995).
5. Houghton, J. T. et al. (eds) *Climate Change 1995: The Science of Climate Change* (Intergovernmental Panel on Climate Change/Cambridge Univ. Press, 1996).
6. Trenberth, K. E. (ed.) *Climate System Modeling* (Cambridge Univ. Press, 1992).
7. Kalnay, E. et al. *Bull. Am. Meteorol. Soc.* 71, 1410–1428 (1990).
8. Lorenz, E. N. *The Essence of Chaos* (Univ. Washington Press, Seattle, 1993).
9. Boer, G. J. et al. *J. Geophys. Res.* 97, 12771–12786 (1992).
10. Cess, R. D. et al. *Science* 267, 496–499 (1995).
11. Kiehl, J. T. et al. NCAR Technical Note NCAR/TN-420+STR (National Center for Atmospheric Research, Boulder, CO, 1996).
12. Kattenberg, A. et al. in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T. et al.) 285–357 (Intergovernmental Panel on Climate Change/Cambridge Univ. Press, 1996).

Electroluminescence in polymer films

Optically pumped lasers, using conjugated polymers as their active material, have been reported both in solution¹⁻³ and in thin films^{4,5}. There is now a great deal of interest in the prospects of producing similar, electrically pumped laser diodes¹⁻⁶. Here we report tunnelling-induced electroluminescence experiments on polymer films which indicate that construction of such devices should be possible.

Microcavity devices⁷ have the appropriate structure for laser diode operation, with in-phase reflection from the cavity mirrors providing the necessary feedback. The role of superfluorescence remains a subject of debate⁶, but more important is the question of whether electrical injection can generate the required excitation densities for lasing to occur. From the threshold optical pump power for the establishment of gain, Hide *et al.*⁴ estimate that a current density in excess of 10^6 A m^{-2} will be needed. This is much higher than is normally sustainable within an electroluminescent light-emitting diode (LED), and some have argued that it is fundamentally inaccessible⁷.

We note that electroluminescence can be generated by tunnelling injection from a scanning tunnelling microscope (STM) tip. In our own ultra-high vacuum experiments with thin films (approximately 3 nm thick) of poly(1,3-phenylene vinylene-co-2,5-dioctyloxy-1,4-phenylene vinylene) on a gold surface, we observe electroluminescence at a tunnelling current of 100 pA for an STM tip bias of -2.5 V relative to the gold anode. Although the electric field at the sharp tip is relatively high, at this tip bias the mean field across the polymer film is approximately 10^7 V cm^{-1} , not significantly greater than used in many organic LEDs⁸. Electrons from the tip can tunnel directly into unoccupied states of the polymer, and on combination with holes injected from the anode, excitons are formed that can decay radiatively with the emission of light.

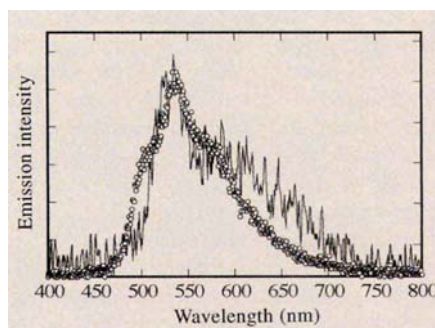


Figure 1 STM-induced electroluminescence emission spectrum (black line) and cathodoluminescence spectrum (circles), in arbitrary units.

The electroluminescent emission was collected and dispersed (Fig. 1), and was stable over periods of at least several minutes (the course of the experiment). The equivalence of the cathodoluminescence spectrum (recorded by raising the STM tip $\sim 1 \mu\text{m}$ above the same spot on the surface and applying a -200 V bias, sufficient to cause field emission from the tip), confirms that the electroluminescence spectrum does indeed arise from the polymer film.

The STM tip was scanned over the surface of the film, and the electroluminescence intensity and topography were simultaneously recorded as a function of position. The smallest features (less than 1 nm for a tunnelling current of 100 pA) resolved in the topographic image provide an upper limit of the tunnelling spot size, an assertion confirmed by the observation of intensity variations in the photon image over similar length scales. Because of the thinness of the film ($\sim 3 \text{ nm}$) and the magnitude of the electric field, we expect little spreading within the film of the current injected through this spot. Hence we conservatively estimate that the applied STM tunnelling current of 100 pA flows through an area of roughly 1 nm^2 , corresponding to a current density of 10^8 A m^{-2} . This exceeds by two orders of

magnitude the estimate in ref. 4. It indicates that there is no intrinsic limit to achieving electroluminescent emission at current densities which (in an optimized device structure) should satisfy the conditions for the establishment of optical gain.

The electroluminescent emission is highly non-uniform on length scales of 10 nm and greater, with variations in both emission intensity and spectral characteristics. Non-uniform emission occurring at similar current densities has also been observed in thin films of poly(*p*-phenylene vinylene)⁹, indicating that these results might apply to a large class of conjugated polymers. This non-uniformity in electroluminescence is problematic for device fabrication, and a detailed understanding of the processes that control local emission efficiency is required as part of any development of electrically pumped polymer lasers. Clearly there is still much to do in engineering a polymer laser diode, but we are convinced that this is an attainable target.

David G. Lidzey

Donal D. C. Bradley

Department of Physics,

University of Sheffield,

Hicks Building, Hounsfield Road,

Sheffield S3 7RH, UK

Santos F. Alvarado

Paul F. Seidler

*IBM Research Laboratory, Säumerstrasse 4,
CH-8803 Rüschlikon, Switzerland*

1. Holzer, W., Penzkofer, A., Gong, S. H., Bradley, D. D. C. & Bleyer, A. *Adv. Mater.* **8**, 974-978 (1996).
2. Moses, D. *Appl. Phys. Lett.* **60**, 3215-3216 (1992).
3. Brouwer, H. J., Krasnikov, V. V., Hilberer, A., Wildman, J. & Hadzioannou, G. *Appl. Phys. Lett.* **66**, 3404-3406 (1995).
4. Hide, F. *et al. Science* **273**, 1833-1836 (1996).
5. Tessler, N., Denton, G. J. & Friend, R. H. *Nature* **382**, 695-697 (1996).
6. Bradley, D. D. C. *Nature* **382**, 671 (1996).
7. Lovinger, A. J. & Rothberg, L. J. *J. Mater. Res.* **11**, 1581-1592 (1996).
8. O'Brien, D., Weaver, M. S., Lidzey, D. G. & Bradley, D. D. C. *Appl. Phys. Lett.* **69**, 881-883 (1996).
9. Alvarado, S. F., Reiss, W., Seidler, P. F. & Strohhriegel, P. *Phys. Rev. B* (submitted).

Adaptation of colour vision to sunlight

The processes of light and dark adaptation in photopic vision are generally thought to be complete within minutes. Even after nearly all the photopigment in the retinal cone cells of the eye has been bleached by exposure to an intense light, both the density of pigment and the observer's sensitivity are said to recover fully in only 7 minutes¹. We describe here a much slower

process, which reveals itself in an alteration of colour vision after an hour's adaptation to English sunlight.

Before and after exposure to natural sunlight, we measured Rayleigh matches, in which red and green spectral lights are mixed to match a monochromatic orange². Using a Nagel anomaloscope, in which two semicircular fields are viewed simultaneously (see Fig. 1a), a subject adjusted the red/green ratio of the upper half-field to match the standard orange field. Data were also obtained with a computer-controlled colorimeter that allowed a 2° red/green

mixture field (690 + 550 nm) to be alternated every 2 s with a monochromatic orange field (590 nm), thresholds being determined with a randomized 'double-staircase' procedure (Fig. 1b). The exit pupils of both instruments were less than 1 mm in diameter and so variations in the natural pupil are unlikely to affect the stimulus.

During the adaptation phase of the experiment the subject read scientific papers illuminated by natural summer sunlight in central Cambridge (typically 80,000 lux). After exposure, Rayleigh matches were shifted in the protan direction — more red

being required in the mixture field. A shift of Rayleigh matches in this direction is known to occur for very short periods after photopigment bleaching^{3,4}, but these previously reported effects last for only seconds or minutes. The present long-term effect, though small (of the order of 2 Nagel units), can be measured reliably for as long as 5 hours after the end of the adaptation period.

In other experiments we have established that the effect shows no measurable transfer between eyes. It can be obtained with an artificial light source (575 W Sirio Daylight Fresnel), giving an illumination of ~40,000 lux at the reading surface; and it survives when the ultraviolet component, or indeed all wavelengths shorter than 460 nm, are removed from this light source.

Because these adapting conditions are

often encountered in the natural world and because matches repeatedly return to pre-adapted levels, we assume that the protan shift (alteration of sensitivity to red light) represents a normal mechanism of adaptation rather than subclinical pathology. The effect is unlikely to occur after the receptor level, because a colour match requires that the standard and matching fields lead to the same triplet of quantum catches in the three classes of retinal cone. Subsequent neural events cannot undo this equation⁵. Therefore, a change must have occurred in either the spectral sensitivity of the cone photoreceptors themselves, or one of the ocular pigments that are usually considered to be photostable.

A change in the spectral sensitivity of the photoreceptors could be caused by a very long-lasting photoproduct, or by a reduc-

tion in the self-screening of the photopigment. The latter could arise either from a reduction in optical density or from a reduction in the alignment of the photoreceptors with the centre of the pupil. We have some evidence against a change in photoreceptor alignment, in that we find a shift in Rayleigh match but no change in the Stiles-Crawford effect⁶ after 30 min of exposure to artificial daylight of 40,000 lux.

Among the ocular pigments conventionally regarded as stable are macular pigment, melanin and lipofuscin. Light exposure might alter the optical density of such a pigment or might prompt a migration of pigment (as has been reported for melanin granules of the pigment epithelium in one mammalian species⁷). By brightness matching experiments in the short-wave region, we have satisfied ourselves that no significant change occurs in macular pigment under the conditions of our experiment; but preliminary measurements suggest that the red/green matching ratio is changed by the same factor for all wavelengths between 570 and 620 nm, a result that does implicate one of the ocular pigments other than the cone pigments themselves.

Rayleigh matches are often said to be bimodally distributed within the normal male population^{8,9}, although we have not found such bimodality¹⁰ (and there is no longer a molecular genetic reason to expect it^{11,12}). The shifts of colour matches reported here are of the same order of magnitude as the individual differences observed when Rayleigh matches are obtained from a group of colour-normal people. Bimodalities might arise artifactually if subsets of subjects differed in their recent history of light exposure (for example by wearing sunglasses or being tested at different times of day). Certainly, the subjects' history of light exposure will need to be controlled in future population studies.

G. Jordan

J. D. Mollon

Department of Experimental Psychology,
University of Cambridge,
Downing Street,
Cambridge CB2 3EB, UK

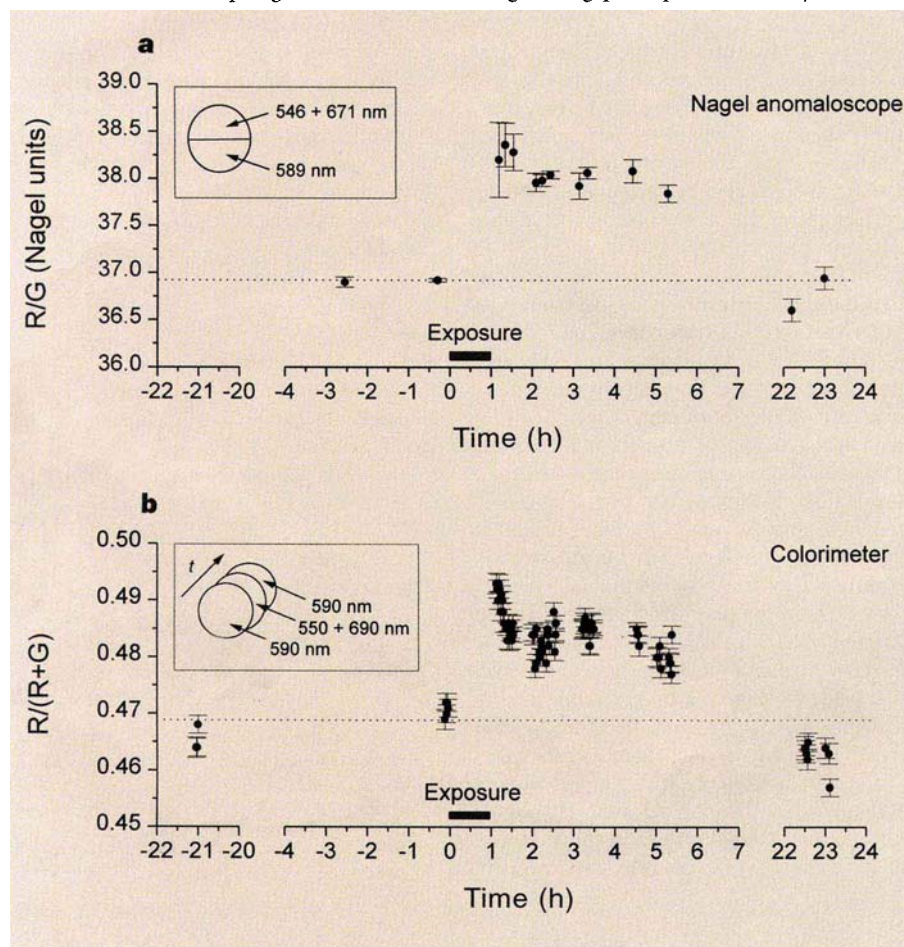


Figure 1 Long-term shift of Rayleigh matches after exposure to 1 h of natural sunlight in June. **a**, Data obtained using a Nagel anomaloscope. The subject (G. J., female, colour-normal) set a match by adjusting the red/green mixture ratio of the upper half-field (546 + 671 nm) and the brightness of the lower orange half-field (589 nm). Room temperature was held at 21 °C (ref. 13). Each data point is the mean $\pm$ s.e.m. of 5 matches. **b**, Data measured for the same subject on a three-channel colorimeter. A temporal substitution method was used (see inset), the alternating mixture and standard fields each having a duration of 2 s. Two 'staircases' were randomly interleaved and on each presentation of the mixture field the subject indicated whether it was too red or too green to match the standard. Each data point is the mean of 20 trials, after which the subject had the opportunity to adjust the brightness of the standard. Both panels show the amount of red in the red/green mixture that matched the orange standard as a function of time. Zero on the abscissa indicates the beginning of the 1-h exposure to sunlight. The dotted line is the mean of all matches before exposure. Similar results have been obtained for a male subject (J. D. M.).

1. Rushton, W. A. H., Fulton, A. B. & Baker, H. D. *Vision Res.* **9**, 1473–1479 (1969).
2. Rayleigh, Lord *Nature* **25**, 64–66 (1881).
3. Wright, W. D. *J. Physiol.* **87**, 23–33 (1936).
4. Brindley, G. S. *J. Physiol.* **122**, 332–350 (1953).
5. Rushton, W. A. H. *J. Physiol.* **220**, 1–31 (1972).
6. Enoch, J. M. & Lakshminarayanan, V. in *Vision and Visual Dysfunction I: Visual Optics and Instrumentation* (ed. Charman, W. N.) 280–309 (Macmillan, Manchester, 1991).
7. Pang, S. F., Yew, D. T. & Tsui, H. W. *Neurosci. Lett.* **10**, 221–224 (1978).
8. Waaler, G. H. M. *Nature* **215**, 406 (1967).
9. Neitz, J. & Jacobs, G. H. *Nature* **323**, 623–625 (1986).
10. Jordan, G. & Mollon, J. D. *Vision Res.* **35**, 613–620 (1995).
11. Asenjo, A. B., Rím, J. & Oprian, D. D. *Neuron* **12**, 1131–1138 (1994).
12. Mollon, J. D. in *Colour Vision Deficiencies XIII* (ed. Cavonius, C. R.) 1–18 (Kluwer, Dordrecht, 1997).
13. Jordan, G. & Mollon, J. D. *Nature* **363**, 546–549 (1993).

Is scrapie solely a genetic disease?

In sheep¹⁻⁶ and goats⁷, polymorphisms of the PrP gene have been strongly associated with incidence of natural and experimental scrapie, the most common transmissible spongiform encephalopathy. However, it is still a matter of some debate whether natural scrapie is a genetic disease, arising from certain sheep PrP gene alleles⁸, or whether PrP controls susceptibility to an infecting agent⁶. Here we present evidence that scrapie is not solely a genetic disease, as scrapie-associated PrP alleles are present in sheep from Australia and New Zealand, both countries that are entirely free of scrapie.

We determined PrP genotypes of healthy adult Cheviot and Suffolk sheep (breeds in which the genetics of scrapie is well understood) from the United Kingdom, Australia and New Zealand. We did not find the genotype at highest risk of contracting scrapie (VV₁₃₆RR₁₅₄QQ₁₇₁; genotype 1 in Fig. 1) among a total of 168 UK Cheviots with ages of 24–96 months, whereas we found genotype 2 (VA₁₃₆RR₁₅₄QQ₁₇₁), also with a high risk of disease, at a frequency of 14% and the lowest risk group, genotype 13 (AA₁₃₆RR₁₅₄RR₁₇₁), was found at 19%.

The Australian Cheviot sample included two sheep (4%; aged 36 and 96 months when sampled) with genotype 1. The PrP gene open reading frames of these two animals were sequenced (not shown), revealing no differences from previously sequenced VV₁₃₆RR₁₅₄QQ₁₇₁ scrapie-affected UK Cheviots. The frequency of genotype 2 was lower in the Australian Cheviots (9%) than in the UK sample, as was the frequency of genotype 13 (9%). The New Zealand Cheviot sample had no genotype 1 sheep among 102 animals, although there was 17% frequency of genotype 2. Genotype 13 was relatively frequent (21%).

Suffolk sheep show much less genetic variation in the PrP gene than Cheviots and have different scrapie-related genotypes^{4,9}. The most susceptible Suffolk genotype, AA₁₃₆RR₁₅₄QQ₁₇₁ (genotype 6), represented 13% of the UK Suffolk sample and the most resistant genotype, AA₁₃₆RR₁₅₄RR₁₇₁ (genotype 13) was present at 33%. In the Australian and New Zealand Suffolk samples, we found genotype 6 at relatively high frequencies (53% and 39%, respectively); genotype 13 was less common (3% and 8%) than in the UK sample.

The measures taken by the Australian and New Zealand authorities to protect sheep from scrapie are lengthy and expensive. In New Zealand, importation of sheep was prohibited from all countries except Australia in the 1950s but, since 1984, sheep

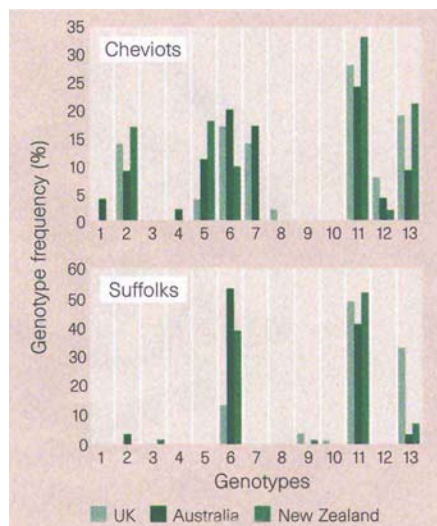


Figure 1 Genotypes were assigned using methods described previously^{3,6,10} and are: 1, VV₁₃₆RR₁₅₄QQ₁₇₁; 2, VA₁₃₆RR₁₅₄QQ₁₇₁; 3, VA₁₃₆RR₁₅₄QH₁₇₁; 4, VA₁₃₆HR₁₅₄QQ₁₇₁; 5, VA₁₃₆RR₁₅₄RQ₁₇₁; 6, AA₁₃₆RR₁₅₄QQ₁₇₁; 7, AA₁₃₆HR₁₅₄QQ₁₇₁; 8, AA₁₃₆HH₁₅₄QQ₁₇₁; 9, AA₁₃₆RR₁₅₄QH₁₇₁; 10, AA₁₃₆RR₁₅₄RH₁₇₁; 11, AA₁₃₆RR₁₅₄RQ₁₇₁; 12, AA₁₃₆HR₁₅₄RQ₁₇₁; 13, AA₁₃₆RR₁₅₄RR₁₇₁. Genotypes 1 and 2 are linked to natural scrapie in Cheviots, whereas genotype 6 is linked to scrapie in Suffolks. The analysis was carried out on 168, 54 and 102 Cheviots and 69, 32 and 112 Suffolks from the United Kingdom, Australia and New Zealand, respectively. There was no link between PrP genotype and age (not shown), which ranged from 12–96 months for Cheviots and 12–108 months for Suffolks.

from countries posing a low risk of carrying scrapie have been introduced. Animals are quarantined for at least three years on an offshore island and, if they remain healthy, their progeny are brought into the national flock by embryo transfer. Any case of scrapie in quarantine (the last occurred in 1977) would result in the slaughter of all animals in the programme (S. MacDiarmid, personal communication). These efforts are worthwhile, as our studies indicate that there are susceptible animals that would succumb to scrapie if the infection was accidentally imported.

Scrapie is, therefore, not a spontaneous genetic disease linked to PrP genotypes alone. Although dependent on PrP genotype, the development of scrapie in sheep must require an additional exogenous factor, most probably an infectious agent that is absent from unaffected flocks and from scrapie-free countries.

N. Hunter, D. Cairns
J. D. Foster, G. Smith
W. Goldmann, K. Donnelly
Institute for Animal Health,
Neuropathogenesis Unit, West Mains Road,
Edinburgh EH9 3JF, UK
e-mail: nora.hunter@bbsrc.ac.uk

- Hunter, N., Foster, J. D., Dickinson, A. G. & Hope, J. *Vet. Rec.* **124**, 364–366 (1989).
- Laplanche, J. L. *et al. Genomics* **15**, 30–37 (1993).
- Hunter, N., Goldmann, W., Smith, G. & Hope, J. *Arch. Virol.* **137**, 171–177 (1994).
- Westaway, D. *et al. Genes Dev.* **8**, 959–969 (1994).
- Belt, P. B. G. M. *et al. J. Gen. Virol.* **76**, 509–517 (1995).
- Hunter, N. *et al. Arch. Virol.* **141**, 809–824 (1996).
- Goldmann, W. *et al. J. Gen. Virol.* **77**, 2885–2891 (1996).
- Ridley, R. M. & Baker, H. F. J. *Gen. Virol.* **77**, 2895–2904 (1996).
- Hunter, N., Moore, L., Hosie, B., Dingwall, W. & Greig, A. *Vet. Rec.* **140**, 59–63 (1997).
- Hunter, N., Goldmann, W., Benson, G., Foster, J. D. & Hope, J. *J. Gen. Virol.* **74**, 1025–1031 (1993).

Assembly of Borromean rings from DNA

Borromean rings are a rich family of topological structures¹, consisting of interlocking rings, whose simplest member (Fig. 1a) appears on the coat of arms of the Borromeo family, prominent in the Italian Renaissance. Their key property is that removal of any individual ring unlinks the remaining rings. Here, we report the assembly of Borromean rings from single-stranded DNA.

Borromean rings have equal numbers of positive and negative crossings (called unit tangles² or nodes) at specific positions. Half-turns of double-helical DNA can provide nodes for topological construction; the sign of each node can be imposed by choosing B-DNA (negative crossings) or Z-DNA (positive crossings)³. We have used this strategy to make trefoil or figure-of-eight knots from DNA⁴ and RNA⁵. A related approach has been used to forge multi-nuclear coordination compounds into trefoil and composite knots⁶.

To make Borromean rings, we have replaced half-turns of DNA with one-and-a-half turns (Fig. 1b), because longer units are more tractable. The key to translating this arrangement into DNA components is to recognize that Fig. 1a and b are topological diagrams, not geometrical representations. They can be considered as similar to polar projections of the Earth, where the North Pole is at the centre (nearest the viewer), and every point on the circumference corresponds to the South Pole (farthest from the viewer).

The left-hand stereo image of Fig. 1c is a three-dimensional representation of Fig. 1b, making it clear that the proper topology is attained by joining a three-arm DNA branched junction⁷ formed from B-DNA (flanking the 'north pole') to a three-arm DNA branched junction formed from Z-DNA (centred on the 'south pole'). They can be joined at the 'equator' (Fig. 1c, right-hand stereo image) by six DNA hairpins. These do not affect the topology of the product and so their size can be varied, allowing the three strands to be of different lengths (196, 206 or 216 residues).

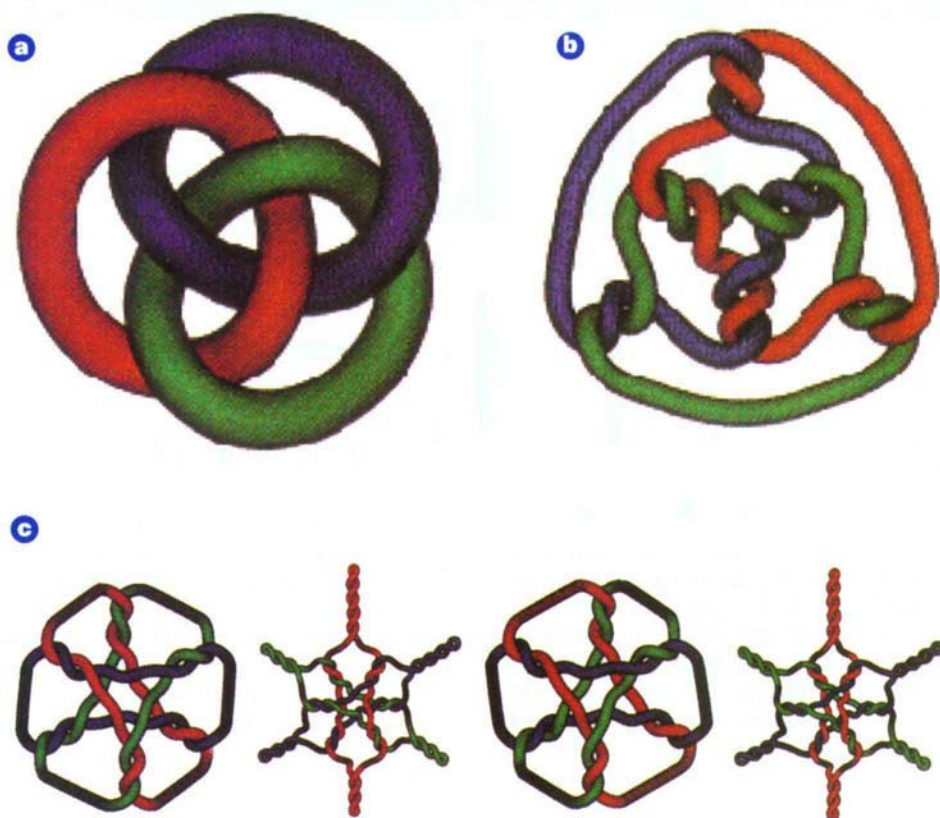


Figure 1 a, Traditional Borromean rings. b, Borromean rings with each node replaced by three nodes, derived from 1.5 turns of double helix. The inner double helices are right-handed and the outer double helices are left-handed. c, Left, stereoscopic representation of (b), where a B-DNA junction flanks the 'north pole' and a Z-DNA junction flanks the 'south pole'. Right, stereoscopic views of the molecules synthesized. Two hairpins have been added to the 'equatorial' sections of each strand. Drawings made using the KnotPlot program, written by R. Scharein; <http://www.cs.ubc.ca/nest/imager/contributions/scharein/KnotPlot.html>

We constructed this DNA Borromean link by synthesizing six strands of DNA corresponding to the three strands of each branched junction. Each junction contains a dT₄ spacer between its arms; the arms of the B-DNA junction contain 16 base pairs, but the Z-DNA junction arms contain 18 base pairs, because the Z-DNA repeat is 12 base pairs⁸. The Z-segments contain 5-methylcytosine, to increase the Z-forming propensity⁹, and occasional A-T pairs for recognition specificity¹. Each junction strand is both preceded and followed by a dT₁₀ spacer that connects it to one of the hairpins: There is an octanucleotide sequence on the 5' end of each strand, and a hairpin on the 3' end, lacking this octa-

nucleotide to allow recognition. Each hairpin contains a site for individual cleavage by a restriction enzyme.

The B-DNA junction and Z-DNA junction were annealed individually, and then combined and ligated under Z-DNA-promoting conditions. The products were separated on two-dimensional (4–6%) denaturing gels⁵, thereby ensuring that the products are both cyclic and covalently linked. The target molecule had the slowest mobility in each direction. Further electrophoresis on 4%, and then 6%, sequencing gels removed traces of an impurity.

The synthesis is confirmed in Fig. 2. Strand 1 contains a site for restriction by *PvuII*, strand 2 for *EcoRV* and strand 3 for

XbaI. Lanes 4–6 contain, respectively, the digestion products from the cleavage of the Borromean rings (top of the gel) by *PvuII*, *EcoRV*, and *XbaI*. In each case, the only cyclic products are the unrestricted individual circles, not catenanes.

By identifying components to serve as positive and negative unit tangles (or their odd multiples), linking them into stable units (here branched junctions), and joining them by recognition-directed ligation we have achieved the construction of Borromean rings from DNA. This approach should provide topological control in other chemical systems. Conversely, it may be possible to use this or other successful systems to act as scaffolding to guide the formation of target topological products from other polymers.

Chengde Mao

Wei Qiong Sun

Nadrian C. Seeman*

Department of Chemistry, New York University,
New York, New York 10003, USA
e-mail: ned.seeman@nyu.edu

1. Liang, C. & Mislow, K. J. *Math. Chem.* **16**, 27–35 (1994).
2. Sumners, D. W. *Math. Intelligencer* **12**, 71–80 (1990).
3. Seeman, N. C. *Mol. Eng.* **2**, 297–307 (1992).
4. Du, S. M., Stollar, B. D. & Seeman, N. C. *J. Am. Chem. Soc.* **117**, 1194–1200 (1995).
5. Wang, H., Di Gate, R. J. & Seeman, N. C. *Proc. Natl Acad. Sci. USA* **93**, 9477–9482 (1996).
6. Carlina, R. F., Dietrich-Buchecker, C. & Sauvage, J.-P. *J. Am. Chem. Soc.* **118**, 9110–9116 (1996).
7. Seeman, N. C. *J. Theor. Biol.* **99**, 237–247 (1982).
8. Wang, A. H.-J. et al. *Nature* **282**, 680–686 (1979).
9. Behe, M. & Felsenfeld, G. *Proc. Natl Acad. Sci. USA* **78**, 1619–1623 (1981).

*To whom correspondence should be addressed.

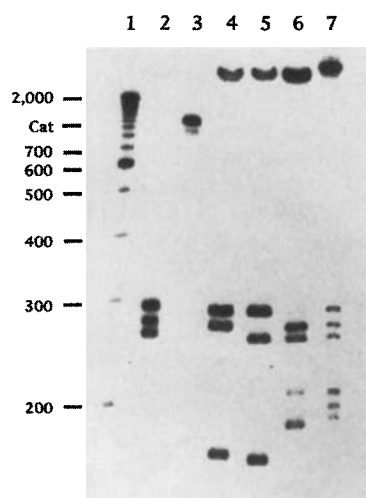


Figure 2 Analysis of the Borromean DNA.

A 6% polyacrylamide denaturing gel containing 40% formamide is shown. Lane 1, 100-base oligonucleotide DNA ladder (GibcoBRL). Lane 2, three individual circles of 196, 206 and 216 nucleotides (C1, C2 and C3). Lane 3, a catenane of circles 2 and 3. Lanes 4–6, the target Borromean rings and the products of their digestion by *PvuII* (strand 1), *EcoRV* (strand 2), and *XbaI* (strand 3), respectively. Lane 7, a mixture of the Borromean rings (top), the three individual circles, and their three linear counterparts (L1, L2 and L3). The scission products of the three rings (shorter than linear molecules) are labelled S1, S2 and S3. Traces of linear strand 3 (most heavily labelled) are visible owing to its spontaneous breakdown. Strand sequences and details of the synthesis methods are available at <http://seemanlab4.chem.nyu.edu>.

The war against cancer

Between Bench and Bedside: Science, Healing, and Interleukin-2 in a Cancer Ward

by Ilana Löwy
*Harvard University Press: 1997. Pp. 370.
\$39.95, £26.50*

Crafting Science: A Sociohistory of the Quest for the Genetics of Cancer

by Joan H. Fujimura
*Harvard University Press: 1997. Pp. 312.
\$45, £29.95*

John Cairns

Cancer is now the cause of about a quarter of all deaths in the developed world. It is lethal because a large proportion of all cancers have spread beyond the reach of surgery when first detected. So there has been an intense search for some non-surgical treatment that, like antibiotics, can find its targets wherever they are. Unfortunately, unlike the best antibiotics, many forms of cancer chemotherapy have proved toxic and dangerous. On the plus side of the ledger, between 5 and 10 per cent of patients whose cancers would have been fatal are now being cured by these drugs; on the minus side, many patients are being exposed to severe treatments that temporarily lower their quality of life and can actually hasten their death.

It must be agonizingly difficult to decide whether to undergo a course of treatment when the results for so many types of cancer are so discouraging. The matter is further complicated, in some countries more than others, by the highly publicized "war against cancer" and the opportunities this has created for fame and fortune. Even the flimsiest claim of success can be the makings of a career, or the basis for a grant application, or the stimulus for a jump in the value of shares. "Would you like us to do all we can?", the relatives are asked, and some doctors scarcely wait for an answer.

In years to come we will no doubt look back on this as an anomaly, a time of frenzy and frustration like those moments of religious crisis scattered throughout past centuries. Historians have a detailed picture of village life in fourteenth-century France because of the Inquisition's investigation of the Cathar heretics in Montailou; future historians of medicine will similarly have to rely on reports from emissaries to the cancer wards of the 1990s. *Between Bench and Bedside* is an important contribution to that literature.

Ten years ago, Ilana Löwy was stimulated by the early enthusiasm for the therapeutic value of cytokines to change the

main emphasis of her career from immunological research to what in effect has been a form of anthropology. She decided to become an observer of the human interactions in a French unit carrying out a trial of interleukin-2 for the treatment of melanomas, renal cancers and neuroblastomas. Her account includes a clear description of the underlying science — which, being immunological, is very complicated. But she is mainly concerned with the way doctors, scientists and patients coped with the gradual realization that interleukin-2 is not all they had hoped it to be.

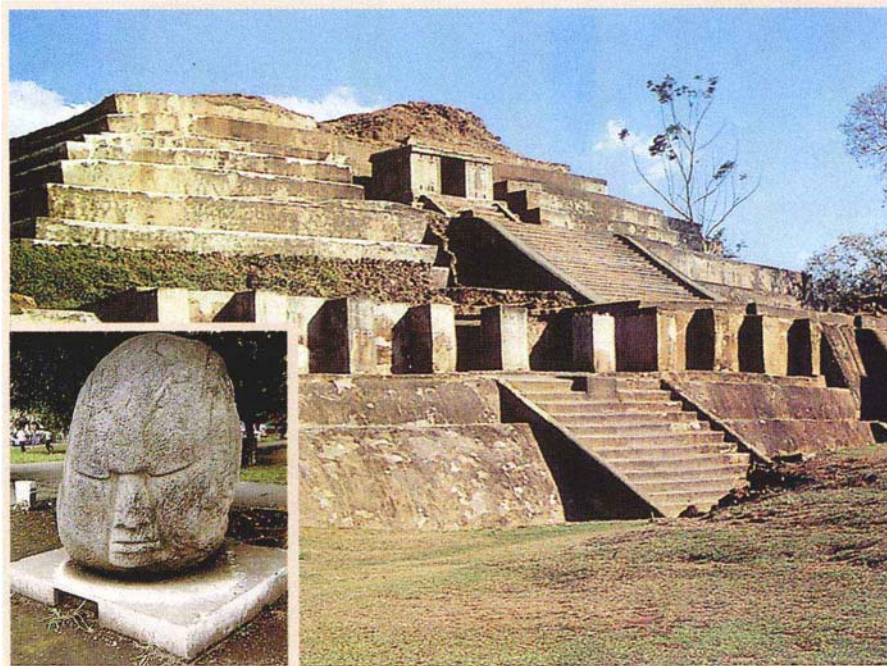
During her years of observation the unit did not, apparently, record any cures, but at least the temporary remissions could be a topic of conversation at staff meetings; the failures and the ultimate fate of the patients after they left the unit were not apparently of much interest. Later, Löwy returns to see how everyone is getting on. Lo!, the unit is being enlarged, the staff have found new agents to enthuse over, and the head of the unit has been promoted.

What makes this book particularly valuable is the way the author keeps herself in

the background. Like a Brontë heroine, she sees all but does not pass judgement. Nowhere did I detect a negative comment about anyone, and the story gains immeasurably from her restraint.

The lack of a certain cure for any of the major forms of cancer, or of a sign that such a cure is on the horizon, suggests that we are going to have to wait until the biological sciences have told us much more about the regulation of cell behaviour. Fortunately, the pace of discovery has greatly accelerated. Thirty years ago, the task of identifying the genes involved in cancer seemed impossibly difficult. Then came retroviruses and oncogenes, and now we have a flood of information. Joan Fujimura's book, *Crafting Science: A Sociohistory of the Quest for the Genetics of Cancer*, is an attempt to do for the discovery of oncogenes what Löwy does for one form of cancer chemotherapy. Fujimura's purpose is presumably to bring about a greater understanding of the way science progresses and how scientists go about their business.

She describes reasonably accurately the developments in cancer research, the dis-



How to get a head in archaeology

One of the most impressive Mayan sites is Tazumal in El Salvador, which is dominated by a large, stepped pyramid (main picture), once crowned with a temple, that dates from the Early to Late Classic period (AD 250 to AD 830). But many earlier Mayan artefacts have been preserved, including the sculptures at Monte Alto in Guatemala (inset), from the

Middle Preclassic period (900 BC to 400 BC).

Those wishing to see the Mayan sites for themselves would do well to consult Joyce Kelly's *An Archaeological Guide to Northern Central America: Belize, Guatemala, Honduras, and El Salvador* (University of Oklahoma Press; \$50 (hbk), \$22.95 (pbk)).

covery of oncogenes, and the techniques for manipulating genes. There are occasional signs that she is not very comfortable with the scientific details, which is to be expected as she is not a practising scientist. But this is not important; anyone who simply wants to learn about oncogenes and genetic engineering would surely choose one of the many books written by scientists. What Fujimura's book has to be judged on is her analysis of the scientific method. And in this respect I have to say that I find her sociohistory utterly unpersuasive.

Part of the difficulty is the matter of style. Scientists search for the simplest way to describe the workings of the world around them. As they learn more, their descriptions usually have to become more complicated, but the art is in making the descriptions as simple as possible. So scientific writing tends to use concrete words of limited meaning that are well known or instantly understandable, rather than abstract words whose meaning may be pushed in almost any direction. In contrast, the writings of social scientists are highly abstract. It is hard to avoid the thought that theirs is a style that satisfies readers who do not want to learn but just to be left with the feeling that information has been paraded in front of them.

For example, Fujimura says about the proto-oncogene theory that scientists "liberally translated the abstract notion to change and add to their laboratories and problem structures without generating major conflicts with preexisting frameworks, yet they simultaneously transmuted their previous notions and technologies". This surely is the language of a culture unaccustomed to the disruptive influence of thought. The style reminds me of the essays I wrote in high school when I had not done the homework and had to fill up pages with obscure verbiage.

Does Fujimura's book serve any purpose? I cannot decide. Too often the meanings of her sentences, when finally worked out, prove to be quite trivial. For example, near the end of a section on what she calls the "Tools of the Trade" she writes: "I seek to understand how scientists come to agreements about how to study nature. From my perspective 'tools', 'jobs', and the 'rightness' of the tools for the jobs are each situationally and interactionally constructed." But the same could be said of the ingredients of almost any purposeful activity — from the stick a chimpanzee pokes into an ants' nest to the supermarket trolley used for the family shopping. Anyone wanting to say something clever about the way scientists operate has to do better than that. □

John Cairns is in the Department of Clinical Trials, Harkness Building, Radcliffe Infirmary, Oxford OX2 6HE, UK.

The comet cometh

Comet of the Century: From Halley to Hale-Bopp

by Fred Schaaf

Copernicus/Springer-Verlag: 1996. Pp. 384. \$29, £21

During the 1985–86 apparition of comet Halley, libraries groaned under the weight of the volumes dedicated to viewing this once-in-a-lifetime event. After Hyakutake last year, and as Hale-Bopp nears perihelion this year, our bookshelves are again under threat from attempts to tell us all about comets (or at least what we know), how to discover them (difficult to say the least), and how Hale-Bopp may be the comet of the century (which it almost certainly won't be).

At least the first book of the onslaught comes from someone who is obviously enamoured of these ghostly objects, rather than someone out to make a quick buck on the tail of a comet. Concentrating on great comets past and present, hence the generic title, Fred Schaaf seems to be on a mission to inspire people to fall in love with comets. This book, written for both the casually interested layperson and the knowledgeable scientist, covers the apparitions of most of the notable comets, with the emphasis placed on those of the nineteenth and twentieth centuries.

As well as the individual comets, Schaaf also recounts the history of cometary research and how the comets themselves have aided and abetted these investigations. A few of the astronomers responsible for the discovery and understanding of comets are in here, but they are rarely more than mentioned in passing. It is the personalities of

the comets themselves, through their changing appearance and behaviour, that have rightly been given top billing as the stars of the show.

As a book on comets in general, it moves between all manner of subjects, from a chronology of apparitions of comet Halley to the now obligatory chapter on collisions of comets with the planets and the Sun. The collision of Shoemaker-Levy 9 with Jupiter in July 1994 certainly adds spice to the proceedings, but it is nice to see that Schaaf does not play up the danger from such events, even (grudgingly) admitting that the collision of the Earth with an asteroid is far more probable than with one of his beloved comets. I was glad to see that this style pervades the book; where a certain hypothesis is unproven or even contentious, such as comet Encke being the progenitor of the hypothetical Taurid stream, Schaaf is generally careful to mention all viewpoints.

Schaaf writes extremely well on the observation and appearance of comets but is a little more shaky when it comes to scientific details. At one point he states that only dust tails shine by reflected sunlight, whereas in fact ion tails can also be observed as a result of scattered solar photons, albeit at discrete wavelengths. As readers of this book are likely to be interested in understanding what they are seeing, such slips are unfortunate, but overall the explanation of the physics and dynamics of comets is sound.

Schaaf has produced a book that is in some ways out of time. The penultimate chapter will certainly be useful in guiding anyone to a view of Hale-Bopp this spring, although I suspect many other books on the subject will do that just as well. The beauty of this book will become apparent only in another few years, when once again we will

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Comet Hale-Bopp can now be seen lighting up our skies.

be in a lull between great comets in the sky. It is at times like those that enthusiasm will need to be rekindled, and these inspiring accounts of bygone comets cannot help but do so. □

Alan Fitzsimmons is in the Department of Pure and Applied Physics, Queen's University of Belfast, Belfast BT7 1NN, UK.

Through a glass, historically

The Microscope in the Dutch Republic: The Shaping of Discovery

by Edward G. Ruestow
Cambridge University Press: 1996. Pp. 348.
£40, \$59.95

G. L'E. Turner

Because of the constant interactions between the instrumental and conceptual requirements of science, with economic demands and technological limitations also to be taken into account, each type of scientific instrument has its own unique process of development. Optical instruments are of particular importance but pose some intractable problems for the science historian. For instance, as glass was made by the Romans, why was it that eye-glasses were not invented until about 1286 in Italy, and why did another 300 years pass before the invention, in 1608, of the telescope, at Middelburg in the Netherlands? However, once the two lenses had been put together to form the optical system of the telescope, a lens system to create a microscope soon followed.

Although the telescope and the microscope were invented almost at the same time, the telescope came quickly into use, and was rapidly established as an essential tool for astronomers. Both instruments suffered from poor-quality glass, so that throughout the eighteenth century the most effective telescopes were not refractors but reflectors that used metal mirrors rather than glass lenses.

The telescope, making the distant heavenly bodies clearly visible to the observer, was much more easily manageable in imaginative and conceptual terms than the microscope because the motion of the planets had been under discussion for thousands of years. The world revealed by the microscope was totally strange, difficult for the observer to understand, and, until the invention of photography, almost impossible to communicate to others. This was particularly true of discoveries about living organisms, for new terminology was needed to supersede the Galenic descriptions familiar to seventeenth-century medical men.

Edward Ruestow takes for his theme the

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Seventeenth-century compound microscopes of Italian design.

first significant discoveries made with the microscope in the century of its invention. His setting is the golden age of Dutch history, when Holland had acquired wealth through sea-going trade and colonization and had achieved cultural pre-eminence in art and learning. His central characters are two Dutchmen, Antoni van Leeuwenhoek (1632–1723) and Jan Swammerdam (1637–80), who used the simple microscope to investigate the natural world, notably the biology of humans and animals. The story is continued in outline through the eighteenth century to the first half of the nineteenth century.

Ruestow offers us an extremely thorough study of a fascinating period in the history of science. He concentrates on the social, philosophical and religious implications, but less so on the technical side. There were two types of microscope, a compound lens version, designed by the Italian academic physician Marcello Malpighi (1628–94), and Leeuwenhoek's simple, or single lens, microscope. Malpighi was the first to overturn the classical theory of the four humours by examining the lungs with a microscope — he was able to show the capillary network that linked arteries and veins, and so to describe the circulation of the blood.

As we now know, the compound version was the eventual winner because it could be mounted with a stage for easier manipulations as well as with a substage illuminating system. But chromatic aberration remained a problem until about 1800 and, although the simple microscope was slightly better in this regard, it was more difficult to use. As

Ruestow explains: "To prepare a small and fragile subject for purposeful study did take time as well as ingenuity, adroitness, and experience; add the challenge of setting the subject before the lens so that the light could reveal what needed to be seen, and the effort could indeed become a trying one."

Ruestow occasionally blames Leeuwenhoek for a lack of encouragement offered to his successors, and for not engaging in discipleship, and points out that Leeuwenhoek and the better-placed Swammerdam kept the microscope out of an institutional setting. This is not really fair to either man. Their work was published, and the academic community could not miss it. Indeed, from 1667, Malpighi delivered his scientific correspondence to the Royal Society of London, and he was elected a fellow in 1669. Similarly, Leeuwenhoek began in 1673 his 50-year correspondence with the Royal Society, and became a fellow in 1680.

Both the French Académie Royale des Sciences and the Royal Society are criticized for not having a commitment to training and so not supporting a research tradition. But the limitation is really to be found in the classical conceptual framework accepted in the medical schools. This framework gradually changed in the early nineteenth century, and, from about 1830, the microscope rapidly became a crucially important scientific instrument. The previously missing standard of measurement was introduced into microscopy, and the required nomenclature was formulated.

This long, dense book must have been many years in the making. It is fully referenced with an extensive bibliography and a good index. Ruestow is to be congratulated for producing a work that will be read with profit for many years to come. □

G.L'E. Turner is visiting professor in the History of Science and Technology Group, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK.

Ribozymes to the fore

Catalytic RNA

Edited by Fritz Eckstein and David M.J. Lilley

Springer: 1996. Pp 417. £116, \$197

Bob Symons

On the tenth anniversary of the *Nucleic Acids and Molecular Biology* series, the editors have decided to concentrate on one theme, RNA catalysis. The story began with the self-splicing of a group I intron, followed by cleavage *in trans* of bacterial tRNA precursors by an unmodified ribozyme, the RNA component of bacterial RNase P. This remains the only

RNA yet identified to act *in trans* as a true ribozyme without prior modification.

The detailed reaction mechanisms and structural manipulations of group I and II self-splicing introns still present many challenges, as discussed here. But there has been considerable progress in other areas, such as the characterization of small, circular, single-stranded pathogenic RNAs of plants. Earlier work indicated that they are replicated *in vivo* by a rolling-circle mechanism in which the circular RNAs are copied to produce multimeric intermediates that are then processed to monomeric units. Attempts to find protein-free, RNA-catalysed processing reactions in some of these plant pathogenic RNAs were successful, and so were born the hammerhead and hairpin ribozymes.

The small size of the hammerhead ribozyme allows its easy manipulation. Its sequence can be split into ribozyme and substrate components so that cleavage could occur *in trans*, and it has been extensively demonstrated that the ribozyme component of the hammerhead can be designed to target and cleave foreign messenger and other RNAs *in vitro*. There are many studies of the

cleavage of foreign RNAs, and on the effort to use the *in trans* reaction *in vivo* for control of intracellular reactions and pathogen infections, but it is still to be determined whether these will be more effective than antisense approaches. Hence it is not surprising that the hammerhead ribozyme dominates this volume, especially as it is also the area of research interest of the two editors.

They have included the three-dimensional crystalline structure of the hammerhead ribozyme, first reported in 1994, and the development of *in vitro* selection techniques in the search for new types of ribozymes. However, we need to ask how many naturally occurring ribozymes are yet to be discovered. Initially the small, circular, plant pathogenic RNAs were a good place to start looking, in view of the discovery of the hammerhead and hairpin ribozymes. Plant viroids are also likely candidates, as two of these, avocado sunblotch viroid and peach latent mosaic viroid, show hammerhead self-cleavage in the plus and minus strands. The lack of success in finding any RNA-catalysed reactions in the other viroids has led some researchers to believe that the processing of multimeric intermedi-

ates is catalysed by protein rather than RNA. I think it is far more likely that ribozymes will be found when we have developed the appropriate experimental approaches to search for them.

This volume is a rich source of information on catalytic RNA, with many of the major players contributing articles. The only obvious omissions are the as yet uncharacterized ribozyme that is found in an 881-nucleotide transcript of a circular *Neurospora* mitochondrial plasmid, and the extensively characterized ribozyme of hepatitis delta virus, the only ribozyme so far identified in animal systems. □

Bob Symons is in the Department of Plant Science, Waite Agricultural Research Institute, University of Adelaide, Glen Osmond, South Australia 5064, Australia.

Spring books supplement

The 27 March issue of *Nature* will contain the Spring Books supplement. Reviewers are likely to include Walter Gratzer on the discovery of the structure of DNA, David Jones (perhaps better known to *Nature* readers as Daedalus) on design, and Sir Crispin Tickell on T. H. Huxley.

Specialist books

Excellent ★★★★★ Good ★★★★ Average ★★★ Poor ★

Relativity and Gravitation

by Philippe Tourrenc

Cambridge University Press: 1996. Pp. 242. £45, \$74.95 (hbk); £16.95, \$29.95 (pbk)

The nature of gravity has intrigued physicists for centuries, and progress in our understanding has required major changes in our view of the world. With a clearly stated view of the philosophical divide between the newtonian and relativistic worlds, Philippe Tourrenc gives a wide-ranging introduction to special and general relativity.

With discussions of applications, general theories and experimental tests, a very broad view is presented. The mathematics associated with general relativity is often a challenge to students, and in the main text the mathematics is kept to a minimum. There is, however, a sound mathematical underpinning provided by the outline derivations of important results that are inserted into the main text.

For undergraduates with a previous knowledge of special relativity, the book provides a route to general relativity, with the chapters on special relativity providing useful acclimatization to the style and philosophy before the general theory is attempted.

Warren Perkins Department of Physics, University of Wales Swansea, UK.

Range	★★★★
Depth	★★
Accuracy	★★★
Up-to-dateness	★★★
Accessibility	★★
Style	★★

Companion to the Cosmos

by John Gribbin

Little Brown: 1996. Pp. 504. \$29.95/Weidenfeld and Nicolson £20

"Cosmos" is just an alternative word for "Universe" but "Companion" gives little hint that Gribbin is presenting us with a 416-page encyclopaedia of modern astronomy and cosmology plus a 67-page chronology of the history of the subject.

Good astronomical encyclopaedias require affirmative answers to three questions: "Is it in?", "Does the entry make sense when you find it?" and "Is the new encyclopaedia so good you can throw away older ones?". Gribbin certainly makes sense. He writes with such admirable clarity and interest that it is difficult to resist the diverting temptation of looking up what he thinks about other things. He concentrates on the exciting ideas and the influential proponents of astronomy. There is bias towards cosmology. There are no entries on constellations or named stars. James Clerk Maxwell gets more lines than Mars. Anthropic principle is in, ephemeris is out.

You will really enjoy this new encyclopaedia. But keep the old ones.

David W. Hughes Department of Physics, University of Sheffield, UK

Range	★★
Depth	★★★★
Accuracy	★★★★
Up-to-dateness	★★★★
Accessibility	★★★★
Style	★★★★

Limb Regeneration

by Panagiotis Tsonis

Cambridge University Press: 1996. Pp. 241. £55, \$74.95

Regeneration of vertebrate body parts is frustratingly limited, but a few amphibians possess quite spectacular powers and can replace an entire limb. An understanding of how this happens would have important medical implications. Limb regeneration probably involves the reawakening of many of the processes used in the development of the embryonic limb. Molecular advances in the past five years have revolutionized knowledge of limb development, and this can now be applied to regeneration.

Unfortunately this book does not give this impression, and possible new directions — exploring the roles of *Sonic hedgehog*, bone morphogenetic proteins and many more molecules now known to be important in limb development — are hardly mentioned. On the other hand, major issues in limb regeneration are set out well, and the review of traditional studies will prove useful to those not in the field.

Konstandina Kostakopoulou and Cheryl Tickle Department of Anatomy and Developmental Biology, University College London, UK

Range	★★
Depth	★★
Accuracy	★★★★
Up-to-dateness	★★
Accessibility	★★★★
Style	★★★

Photonic crystals: putting a new twist on light

J. D. Joannopoulos, Pierre R. Villeneuve & Shanhui Fan

Photonic crystals are materials patterned with a periodicity in dielectric constant, which can create a range of 'forbidden' frequencies called a photonic bandgap. Photons with energies lying in the bandgap cannot propagate through the medium. This provides the opportunity to shape and mould the flow of light for photonic information technology.

For the past 50 years, semiconductor technology has played a role in almost every aspect of our daily lives. The drive towards miniaturization and high-speed performance of integrated electronic circuits has stimulated considerable research effort around the world. Unfortunately, miniaturization results in circuits with increased resistance and higher levels of power dissipation, and higher speeds lead to a greater sensitivity to signal synchronization. In an effort to further the progress of high-density integration and system performance, scientists are now turning to light instead of electrons as the information carrier.

Light has several advantages over the electron. It can travel in a dielectric material at much greater speeds than an electron in a metallic wire. Light can also carry a larger amount of information per second. The bandwidth of dielectric materials is significantly larger than that of metals: the bandwidth of fibre-optic communication systems is typically of the order of one terahertz, while that of electronic systems (such as the telephone) is only a few hundred kilohertz. Furthermore, light particles (or photons) are not as strongly interacting as electrons, which helps reduce energy losses.

In spite of the numerous advantages of photons, all-optical circuits have yet to be commercially available on a large scale. Some hybrid optoelectronic circuits have produced significant improvement over the performance of electronic circuits, but the difficulties in designing a multipurpose optical component analogous to the electronic transistor has severely hindered the proliferation of all-optical systems. A new class of optical materials known as photonic crystals¹ may hold the key to the continued progress towards all-optical integrated circuits. Indeed, scientists have begun imagining photonic integrated circuits which resemble microscopic metropolises at micrometre length scales, with photonic crystal buildings that house bundles of light, and highways and bridges that guide light along narrow channels and around tight corners.

The underlying concept behind photonic-crystal materials stems from early ideas by Yablonovitch² and John³. In a nutshell, the idea is to design materials so that they can affect the properties of photons, in much the same way that ordinary semiconductor crystals affect the properties of electrons. Both Yablonovitch and John suggested that structures with periodic variations in dielectric constant could influence the nature of photonic modes in a material; Yablonovitch's aim was to control the radiative properties of materials, while John's was to effect photon localization^{4–7} by introducing a random refractive-index variation.

Traditionally, the manipulation of optical photons has relied in general on the mechanism of total internal reflection. Light propagating in a high-dielectric material is reflected at the interface with a low-dielectric material. This severely limits the degree of miniaturization of optical components because the interface must be smooth with respect to the wavelength of light. Photonic crystals offer a completely different mechanism for the control of light. The

difference lies in the concept of a photonic bandgap—the optical analogue of the electronic bandgap in semiconductors.

In a semiconductor, the atomic lattice presents a periodic potential to an electron propagating through the electronic crystal. Moreover, the geometry of the lattice and the strength of the potential are such that, owing to Bragg-like diffraction from the atoms, a gap in allowed energies opens up for which an electron is forbidden to propagate in any direction. In a photonic crystal, the periodic 'potential' is due to a lattice of macroscopic dielectric media instead of atoms. If the dielectric constants of the constituent media are different enough, Bragg scattering off the dielectric interfaces can produce many of the same phenomena for photons as the atomic potential does for electrons. Thus a photonic crystal could be designed to possess a complete photonic bandgap—a range of frequencies for which light is forbidden to exist within the interior of the crystal. Forbidden, that is, unless there is a defect in the otherwise perfect crystal. A defect, or mistake in the periodicity, could lead to localized photonic states in the gap, whose shapes and properties would be dictated by the nature of the defect. A point defect could act like a microcavity, a line defect like a waveguide, and a planar defect like a perfect mirror. This ability to manipulate a photon provides us with a new dimension in our ability to mould or control the properties of light. Therein lies the exciting potential of photonic crystals.

Of course, while the periodic arrangement of atoms occurs naturally in semiconductors, photonic crystals need to be fabricated artificially. To fully appreciate the challenge of fabricating these structures, we note that the lattice constant of the photonic crystal (that is, the size of the basic unit cell) must be comparable to the wavelength of the light. For the optoelectronics industry, for instance, where the usual operating frequency is around 1.5 μm (in the infrared), the lattice constant of a useful photonic crystal must be of the order of 0.5 μm . Although this is about 1,000 times larger than the lattice constant of atomic crystals, it is still over 100 times smaller than the diameter of a human hair. The fabrication of these intricate structures requires state-of-the-art micro-lithography techniques, such as electron-beam lithography and X-ray lithography.

From a theoretical point of view, the description of light in photonic crystals must involve the solution of Maxwell's equations in a periodic dielectric medium. An appealing aspect of Maxwell's equations is that, unlike the complex strongly interacting many-particle problem of electrons in a solid, they can be solved exactly. With linear materials there are no interactions between photons so that one is left with a fairly standard single-particle problem. This means that theoretical computations can provide very accurate descriptions and predictions of the properties of photons and therefore be very useful and complementary to experimental investigations. Another appealing aspect of Maxwell's equations is

that there is no fundamental length scale. If we ignore changes in the dielectric function with frequency, a photonic crystal designed at one length scale will have the same fractional gap as the crystal at any other length scale. Thus a given photonic crystal designed to operate at microwave frequencies, which is much easier to fabricate because of millimetre-scale feature sizes, can be used to deduce the properties of the same photonic crystal scaled down to submicrometre length scales.

To take full advantage of the potential of photonic crystals, it is necessary to develop an understanding and intuition of their photonic structure on the same level as we have for electrons in solids. Towards this end, it is natural to begin our discussion with an exploration of the physical origins of the photonic bandgap. This will illustrate how different dielectric lattice topologies can lead to gaps for different polarizations of light. We shall then proceed to investigate the photonic structure associated with line defects (waveguides) and point defects (microcavities). Waveguides in photonic crystals will provide a unique ability for guiding optical light in air, along narrow channels and around very tight bends, with no losses. Microcavities in photonic crystals will provide complete tunability in both defect frequency and symmetry; the latter leads to the new concept of an *orbital* angular momentum for the photon. Both will provide basic ingredients for using photonic crystals to control the spontaneous emission of atoms in materials. For simplicity, in all of these discussions, we shall concentrate on employing two-dimensionally periodic photonic crystals as generic examples. This is purely for ease of description and visualization. All of the conclusions drawn carry over to their more complex three-dimensional counterparts. While much of this discussion is based on theoretical modelling, several approaches have now been initiated to design and fabricate real three-dimensional photonic crystals.

The photonic bandgap

In the absence of external currents and sources, Maxwell's equations can be cast in the following simple form

$$\left\{ \nabla \times \frac{1}{\epsilon(\mathbf{r})} \nabla \times \right\} \mathbf{H}(\mathbf{r}) = \frac{\omega^2}{c^2} \mathbf{H}(\mathbf{r}) \quad (1)$$

where $\mathbf{H}(\mathbf{r})$ is the magnetic field of the photon, ω is its frequency, c is the speed of light and $\epsilon(\mathbf{r})$ is the macroscopic dielectric function. The solutions $\mathbf{H}(\mathbf{r})$ and ω are determined completely by the strength and symmetry properties of $\epsilon(\mathbf{r})$. If $\epsilon(\mathbf{r})$ is perfectly periodic, as in a perfect photonic crystal, the solutions are characterized by a wavevector $\mathbf{k}$ and a band index n . The region of all allowed wavevectors is called a Brillouin zone and the collection of all solutions is termed a band structure. We are interested in understanding the aspects of the dielectric geometry of a photonic crystal which lead to a complete bandgap in this band structure, that is, a region of frequencies with no allowed photon modes for any value of the wavevector $\mathbf{k}$ inside the Brillouin zone. For simplicity, we examine the band structure of a two-dimensionally periodic photonic crystal comprised of a square lattice of dielectric rods surrounded by air. This system is convenient from points of view of ease of theoretical calculation and ease of experimental fabrication and measurement.

In the top panel of Fig. 1 we illustrate a comparison between experiment⁸ and theory⁹ for the dispersion relations $\omega_n(\mathbf{k})$ of photons in a square lattice of alumina rods ($\epsilon = 8.9$). For the measurements, rods of diameter 0.74 mm and length 100 mm were arranged in a square array, as indicated in the inset, with a lattice constant $a = 1.87$ mm. Coherent microwave transient spectroscopy measurements were then performed for wavevectors $\mathbf{k}$ from (0,0) to $(\pi/a, 0)$, or equivalently from Γ to X, to measure the corresponding frequencies of the propagating photon. Because of the presence of a mirror symmetry plane perpendicular to the rods,

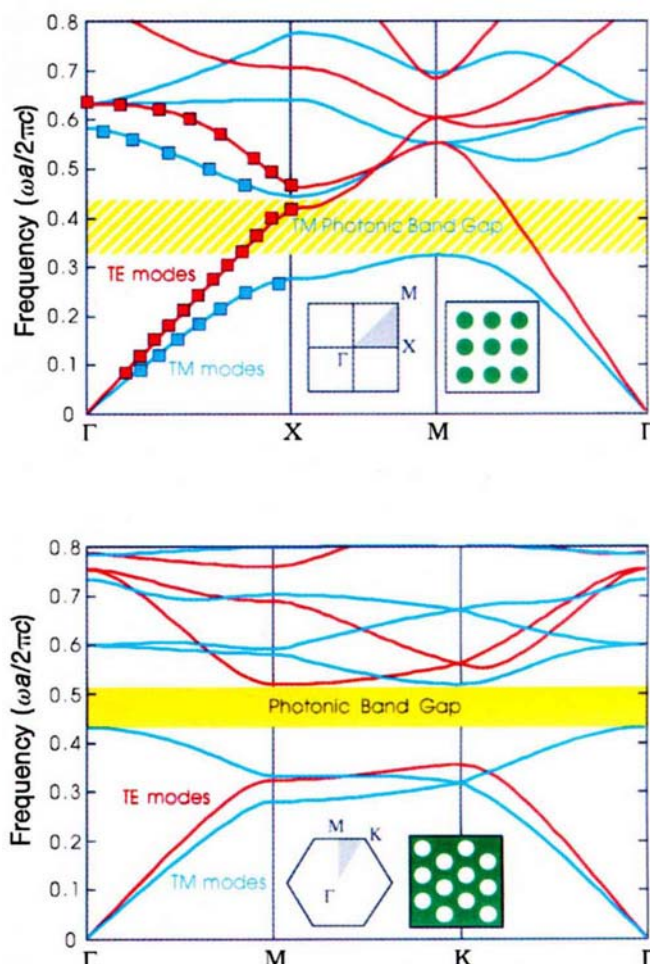


Figure 1 Top, photonic band structure for a square lattice of dielectric ($\epsilon = 8.9$) rods in air with radius $r = 0.2a$, where a is the lattice constant. TM modes are shown in blue and TE modes in red. The solid lines are from theory and the squares represent experimental measurements along Γ to X from Robertson *et al.*⁸ Bottom, photonic band structure for a triangular lattice of air cylinders ($r = 0.48a$) in dielectric ($\epsilon = 13$). Note the presence of a complete photonic bandgap for both TE and TM polarizations in this case as shown by a solid yellow bar. In both panels high-dielectric material is indicated in green in the insets.

it is possible and convenient to decouple the photon modes into polarizations with transverse magnetic (TM) and transverse electric (TE) fields with respect to the plane normal.

The agreement between experiment and theory is excellent for both the TM and TE modes. We note also that for the TM modes there is an experimental indication of a large photonic bandgap between the first and second bands. This is substantiated by the calculation of the bands for the other high-symmetry directions of the Brillouin zone, as shown in Fig. 1. A complete bandgap does indeed exist between the first and the second TM bands. There is, however, no corresponding bandgap for the TE modes. It should be possible to explain such a significant fact.

If we examine the displacement field pattern associated with the lowest TM band we find that it is strongly concentrated in the dielectric regions. This is in sharp contrast to the field pattern associated with the second TM band which has most of its energy in the air regions. These statements have been quantified by calculation⁹ of the fraction, f , of electrical energy inside the dielectric regions. For the modes at the X-point, for example, one obtains $f = 0.8$ and $f = 0.3$ for bands 1 and 2, respectively. The first band has most of its power in the dielectric regions and has a low

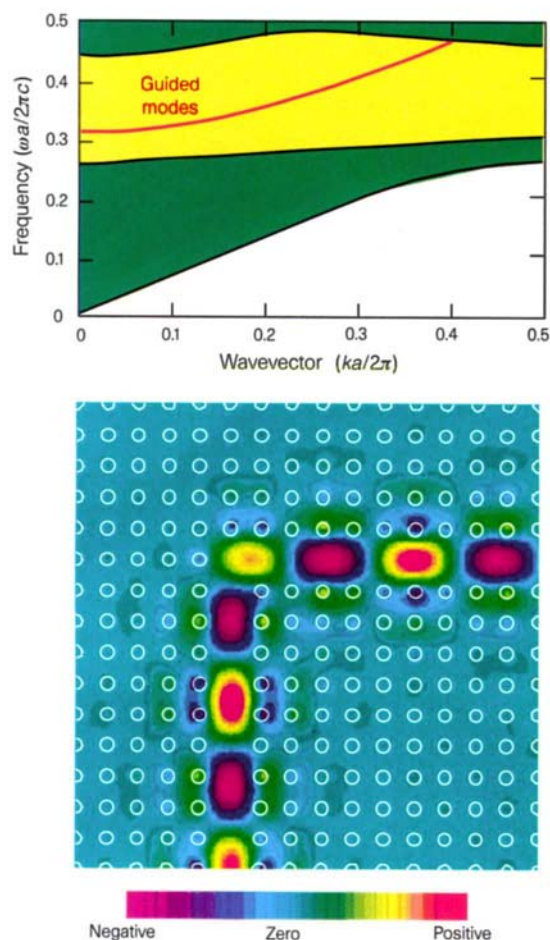


Figure 2 Top, projected photon bands for a waveguide in a square lattice of dielectric rods. The waveguide is formed by removing one row of rods from the otherwise perfect lattice and this creates a band of guided modes as shown in red. The green regions correspond to propagating modes in the bulk crystal and the yellow region corresponds to the photonic bandgap. Bottom, electric field of light propagating down a waveguide with a sharp bend carved out of a square lattice of dielectric rods. The white circles indicate the positions of the rods.

frequency; the second has most of its power in the air region, and has a much higher frequency.

The fractions f for the TE modes do not contrast as strongly. At the X-point, for example, one finds $f = 0.2$ and $f = 0.1$ for the first and second bands, respectively. In this case, both modes have significant amplitude in the air regions, raising their frequencies. They have no other choice; the field lines must be continuous so they are forced to penetrate the air regions. This is the origin of the small values of f and explains the absence of a bandgap for the TE modes. Note that the vector nature of the photon field is central to this argument. The scalar displacement field along the rods of the TM modes can be localized within the rods, but the continuous field lines of the TE modes are compelled to penetrate the air regions to connect neighbouring rods. As a result, consecutive TE modes do not exhibit markedly different f factors, and bandgaps do not appear.

Although we will not discuss it any further here, it is interesting to note that exactly the opposite behaviour is found for TE and TM modes in the case of a crystal with a connected dielectric lattice: an in-depth discussion of this and other aspects of the nature of the photonic bandgap is given in ref. 9. We shall only state the general

rule of thumb: TM band gaps are favoured in a lattice of isolated high- ϵ regions, and TE gaps are favoured in a connected lattice.

This rule of thumb may be used to design a photonic crystal that has a gap for both TE and TM modes. The approach is a sort of compromise: crystals with high- ϵ regions that are both practically isolated and linked by narrow veins. An example of such a system is the triangular lattice of air columns shown in the bottom panel of Fig. 1. A complete photonic bandgap clearly exists for both TE and TM polarizations^{10,11}. Such a gap has recently been observed experimentally in the near-infrared regime along the Γ to K and Γ to M directions for deeply-etched bulk structures¹² and for thin integrated waveguide structures¹³.

It is also possible to find photonic-crystal structures which are not connected, yet exhibit a complete photonic bandgap in the higher-lying bands. This can occur when there is more than one 'dielectric-atom' per lattice constant. An example is the honeycomb lattice of dielectric rods^{11,14}. These types of crystal have the advantage that the larger value of the mid-gap frequency results in a larger value of the minimum feature size. This can be a very important issue in fabrication.

It turns out to be quite typical that the bands above and below a bandgap can be distinguished by where the power lies—in the high- ϵ regions, or in the low- ϵ (usually air) regions. For this reason it is convenient to refer to the band above a photonic bandgap as the 'air band' and the band below a gap as the 'dielectric band'. This is in direct analogy with the use of the terms 'conduction band' and 'valence band' for the electronic band structure of semiconductors.

The waveguide

Once we have a photonic crystal with a gap we can introduce a defect to attempt to trap or localize the light. If we use a line defect, we can also guide light from one location to another. The basic idea is to carve a waveguide out of an otherwise-perfect photonic crystal. Light that propagates in the waveguide with a frequency within the bandgap of the crystal is confined to, and can be directed along, the waveguide. This is a truly novel mechanism for the guiding of light. Traditionally, optical light can be guided without losses within dielectric waveguides such as fibre-optic cables, which rely exclusively on total internal reflection. But, if a fibre optic curves tightly, the angle of incidence is too large for total internal reflection to occur, so light escapes at the corners and is lost. Photonic crystals, on the other hand, continue to confine light even around tight corners.

To illustrate these ideas, we turn again to the square lattice of dielectric rods as a simple example and consider only the TM modes. In the top panel of Fig. 2 we plot the projected bands along the direction of propagation for a waveguide formed by removing one row of rods. The green regions correspond to states that can propagate through the crystal. The band of states within the gap region corresponds to guided modes, which can travel freely within the narrow waveguide channel.

Once light is induced to travel along the waveguide it really has nowhere else to go. An intriguing aspect of photonic-crystal waveguides is that they provide the means to guide optical light, tractably and efficiently, through narrow channels of air. As the frequency of the guided mode lies within the photonic bandgap, the mode is forbidden to escape into the crystal. The primary source of loss can only be reflection back out of the waveguide entrance. This suggests that we may use a photonic crystal to guide light around tight corners. This is also shown in the bottom panel of Fig. 2, as obtained from transmission simulations by Mekis *et al.*¹⁵. Even though the radius of curvature of the 90° bend is zero, 98% of the power in the light that goes in one end comes out of the other. This should be contrasted with 30% power transmission in an analogous dielectric waveguide¹⁵. Making very sharp lossless bends in waveguides which operate at 1.5 μm is of great practical importance for enabling miniaturization of optoelectronic components and circuits.

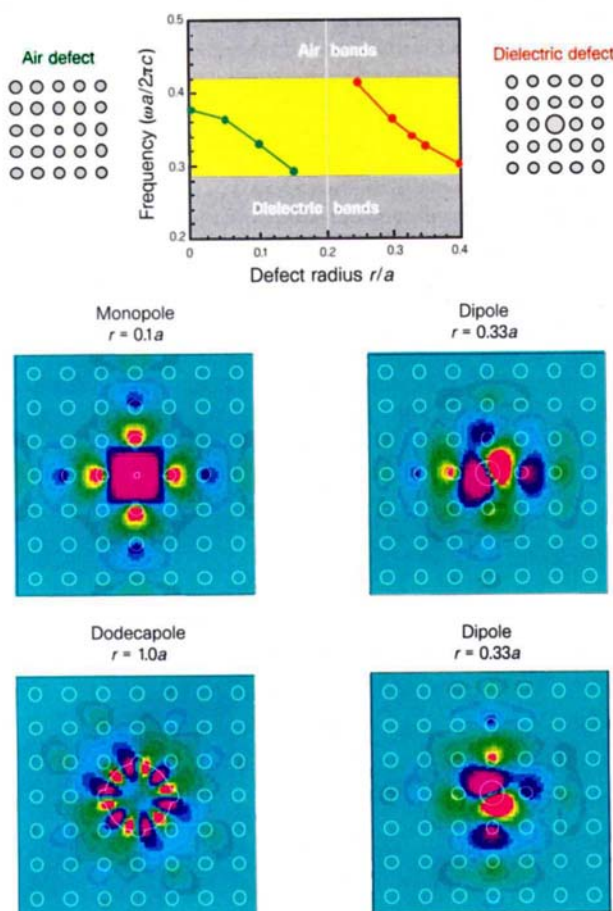


Figure 3 Top, localized states in the gap for a defect formed by varying the radius r of a single rod in the square lattice of dielectric rods with lattice constant a . The case where there is no defect corresponds to $r = 0.2a$, while the case of a vacancy corresponds to $r = 0$. Bottom panels, electric field patterns associated with selected defect states as indicated. The colour coding is the same as in the bottom panel of Fig. 2.

The cavity

One can also create imperfections to trap light at a point within the crystal. One class of imperfections of this type involves changing the dielectric medium in some local region of the crystal, deep within its bulk. As a simple example, consider making a change to a single 'dielectric atom' by modifying its dielectric constant, modifying its size, or simply removing it from the crystal. The top panel of Fig. 3 shows the consequence of creating a vacancy in the square lattice of rods (a vacancy corresponds to a defect radius of $r = 0$). A defect state does indeed appear in the photonic bandgap leading to a strongly localized state. By removing a rod from the lattice, we effectively create a cavity which is surrounded by reflecting walls. If the cavity has the proper size to support a mode in the bandgap, then light cannot escape, and we can pin the mode to the defect.

If the defect involves removal of dielectric (an 'air defect' as in the case of the vacancy) then the cavity mode evolves from the dielectric band and can be made to sweep across the gap by adjusting the amount of dielectric removed. Similarly, if the defect involves the addition of extra dielectric material (a 'dielectric defect') then the cavity mode drops from the air band. As shown in the top panel of Fig. 3, in both cases the defect state can be tuned to lie anywhere in the gap.

Apart from tuning the frequency, one also has some control over the symmetry of the localized photonic state. For example, the

middle and bottom panels of Fig. 3 show the symmetries of the localized photon mode for three different values of the defect radius. For the case of $r = 1.0a$ we find a field pattern that is very reminiscent of the 'whispering gallery' mode observed in microdisk laser cavities¹⁶.

The very specific symmetry associated with each photon mode translates into an orbital angular momentum for each photon mode which can exist in addition to its intrinsic spin angular momentum. This is a very intriguing notion that can have spectacular consequences in the selection rules of electronic transition rates, as discussed below. Further information about air and dielectric defects can be found in refs 17–25.

The flexibility in tuning the symmetry, frequency and localization properties of defects makes photonic crystals a very attractive medium for the design of novel types of filters, couplers, lasers, light-emitting diodes (LEDs) and so on^{16,26}. In the case of laser or LED cavities, photonic crystals provide a particularly unique capability—the control of spontaneous emission.

Spontaneous emission is the natural tendency for an excited atom to 'fall' to a state of lower energy while releasing its energy in the form of emitted radiation. This process, which occurs independently for each atom in a crystal, is at the heart of every light-emitting device used in the optoelectronic industry. LEDs, for example, emit light from the radiative recombination of electrons and holes in a forward-biased p–n junction. Moreover, by increasing the applied voltage, the number of electron–hole pairs in the junction region can become sufficiently large for stimulated emission (that is, the emission induced by other photons) to become prevalent. The junction may then be used as a diode laser.

The ability to control spontaneous emission could have profound consequences for many optoelectronic devices. The rate at which atoms decay depends on the coupling between the atom and the photon, and also on the density of electromagnetic modes available for the emitted photon. Photonic crystals could be used to control each of these two elements independently, simply by changing the properties of the defect states. For instance, the coupling between the atom and the photon involves an integral over all space of the initial and final states of the atom, and of the vector-potential associated with the photon. In the usual dipole approximation, one assumes a uniform vector-potential in the vicinity of each atom, which leads to the standard 'selection rules' for atomic transitions as illustrated in the left panel of Fig. 4. But in the case of a photonic crystal, we can engineer the vector-potential to have specific orbital symmetry. By doing so, transitions that are usually allowed could be made forbidden, and more interestingly, electronic transitions that were previously forbidden could now be made allowed, by judiciously choosing the orbital angular momentum of the defect-state photon, as illustrated in the right panel of Fig. 4. This latter case would be possible if the wavelength of the electron and that of the photon were designed to be of the same order. A synergy between quantum-well and photonic-crystal technologies might be one approach for fabricating systems for which this phenomenon could be measured experimentally.

In addition to changing the coupling between the atom and the photon, the rate of spontaneous emission could also be affected by changing the density of allowed states. If we assume a non-zero coupling between the atom and the photon, the 'natural' rate of emission, in free space, is proportional to the free-photon density of states per unit volume, D_f , which scales as

$$D_f \approx \frac{1}{\omega} \frac{1}{\lambda^3} \quad (2)$$

where ω is the frequency of the transition and λ the wavelength of light. If the system is a photonic crystal with a photonic bandgap around ω , there are no allowed modes to couple to and spontaneous emission is severely inhibited. Conversely, if the photonic crystal is

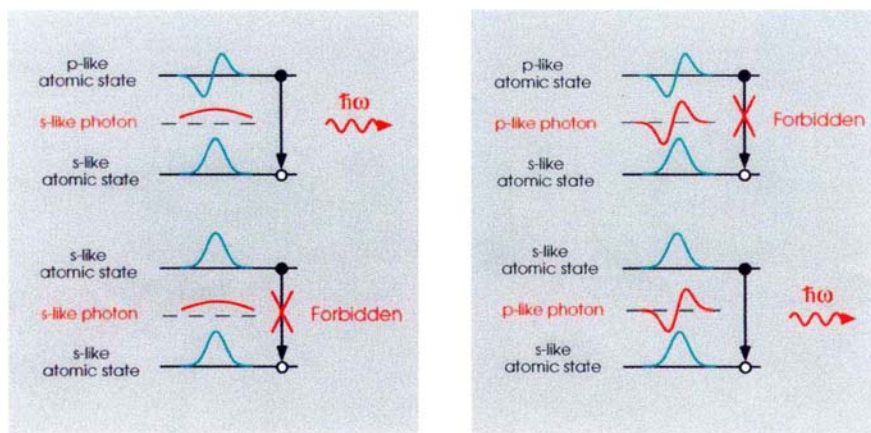


Figure 4 Schematic representation of allowed and forbidden optical transitions for an electron between two atomic states. The symmetry of the orbital angular momentum for the atomic states is shown in blue; the symmetry of the orbital

angular momentum of the photon state is shown in red. The panel on the left corresponds to the usual 'selection rules' for atomic transitions. The panel on the right indicates novel selection rules possible using photonic crystals.

designed to have a point defect with a localized, or more generally, a resonant state at ω , then the emission rate could be enhanced dramatically by the increase in the density of final states. An estimate of this enhancement can be obtained from the following simple argument. The density of states per unit volume for the resonance, D_r , will scale as

$$D_r \approx \frac{1}{\Delta\omega} \frac{1}{\Omega} \quad (3)$$

where $\Delta\omega$ is the frequency width of the resonance and Ω is its effective spatial volume. The enhancement factor is then given roughly by

$$\frac{D_r}{D_f} \approx \frac{\omega}{\Delta\omega} \frac{\lambda^3}{\Omega} = \frac{Q}{(\Omega/\lambda^3)} \quad (4)$$

where $Q \equiv (\omega/\Delta\omega)$ is the quality factor of the cavity. Thus high Q and small spatial volumes can lead to significant enhancement of spontaneous emission. As the smallest volume Ω must be of the order of λ^3 , the largest enhancement will be of the order of Q . Cavities whose spatial volume is of the order of the wavelength of light are called microcavities.

Experimentalists have begun exploring the possibilities of fabricating these microcavities with silicon-based materials and with III-V semiconductor-based materials at micrometre and submicrometre length scales. An exciting new design of a microcavity involves an integrated waveguide-microcavity configuration²⁷, two embodiments of which have recently been successfully fabricated, as shown in the scanning electron micrographs of Fig. 5 (J. Forezi, L. Kimerling and H. Smith, and K. Lim, G. Petrich and L. Kolodziejski, unpublished results). The basic design employs a one-dimensional photonic crystal to confine light in the direction along the waveguide and near the defect at the centre, and total internal reflection to confine light in the transverse directions. The structure in the top panel of Fig. 5 corresponds to a monorail-like geometry fabricated using Si and SiO₂ with electron-beam lithography. The structure in the bottom panel is a novel suspended air-bridge geometry fabricated using GaAs and AlGaAs with a photolithographic technique. In each case, the defect at the centre sustains only one cavity mode whose field patterns are localized to within half a wavelength along the guide and decay rapidly into the air regions. Calculations of the Q for such cavity modes, even for non-optimized geometries, are as high as 10^4 (refs 22, 28). Moreover, and indeed very importantly, we expect the properties of such cavities to be rather robust with respect to random defects. Such disorder will

typically arise during fabrication, and the amount of disorder will of course depend on the minimum feature sizes and the fabrication technique. Calculations²⁹ for a structure with surface disorder whose average size is as large as $\pm 10\%$ of the width of the waveguide, reveal that the Q drops by only 30%. There are two basic reasons for this. First, the wavelength of the mode is significantly larger than the characteristic size of the defects. Second,

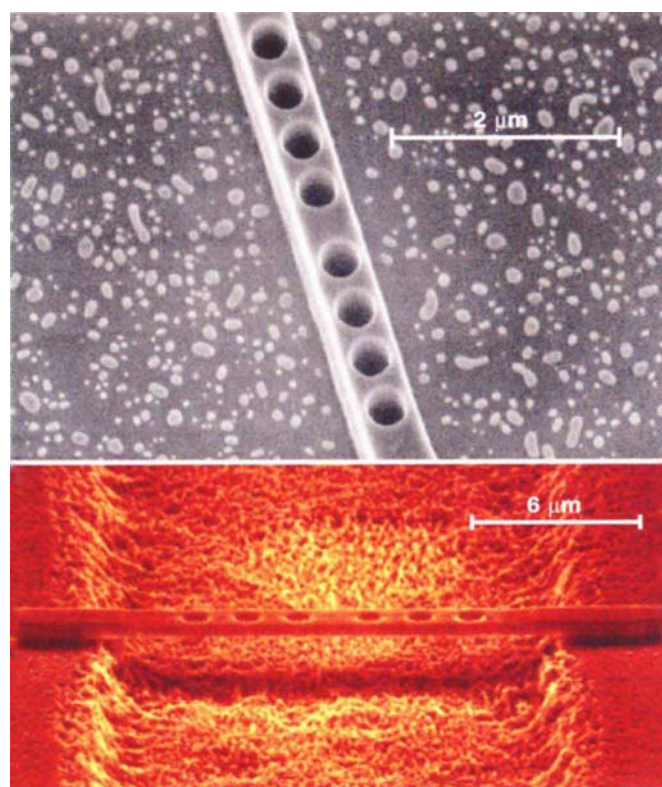


Figure 5 Scanning electron micrographs of two embodiments of a waveguide-integrated photonic crystal microcavity. Top, the monorail geometry fabricated using Si on SiO₂ with electron-beam lithography (J. Forezi, L. Kimerling and H. Smith, unpublished results). Bottom, the air-bridge geometry fabricated using GaAs on AlGaAs with a photolithographic technique (K. Lim, G. Petrich and L. Kolodziejski, unpublished results).

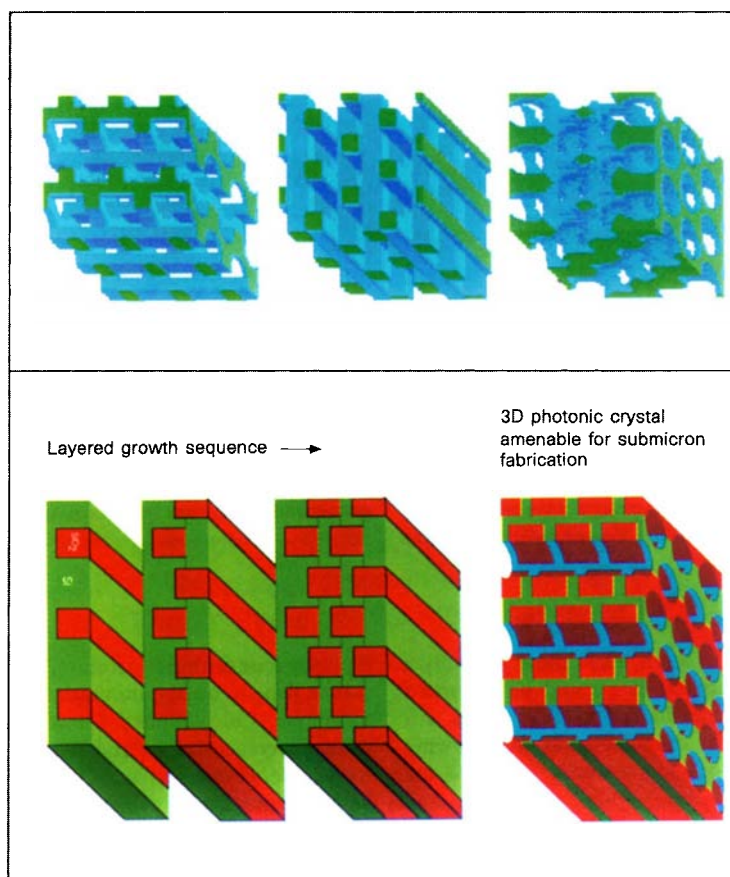


Figure 6 Photonic crystal structures with 3D periodicity (see text for details). Left panel: top, 'Yablonovite' (ref. 30); middle, structure developed by Ho *et al.* (ref. 33); bottom, structure developed by Fan *et al.* (ref. 37). Right panel: lower three

structures, schematic growth sequence for the Fan *et al.* structure; top, final structure obtained after etching vertical holes.

most of the energy of the mode is concentrated in the middle of the cavity and away from the surface. This makes the effect of the surface roughness much less significant for this structure than, say, for microdisks in which the high-*Q* modes propagate along the boundary of the disks¹⁶.

The three-dimensional photonic crystal

The first three-dimensionally (3D) periodic photonic crystal possessing a complete bandgap was fabricated by Yablonovitch³⁰ in 1991 for the microwave regime. The fabrication technique involved covering a slab of dielectric with a mask consisting of a triangular array of holes. Each hole was then drilled three times along (110)-type directions. The resulting structure (known affectionately as 'Yablonovite') is shown at the top of the left panel in Fig. 6. Since then, several other 3D photonic-crystal designs have appeared that offer complete photonic bandgaps^{31–34}.

Of these, the Ho *et al.* structure³³ (shown in the middle of the left panel of Fig. 6) is the smallest 3D photonic crystal with an experimentally demonstrated complete bandgap to be manufactured to date. Özbay *et al.*³⁵ have used a technique of stacking thin micromachined (110) silicon wafers to fabricate these photonic crystals for wavelengths approaching 600 μm .

The ultimate goal for optoelectronic applications is to design and fabricate 3D photonic crystals at an operating wavelength of 1.5 μm . This is certainly not a trivial task. Most recently, Scherer and co-workers³⁶ performed a set of experiments to demonstrate the exciting possibility of fabricating 'Yablonovite' at 1.5 μm , using electron-beam lithography to 'drill' the (110)-type air channels. The first several layers of 'Yablonovite' have been created in this fashion.

A new class of photonic crystals, designed specifically to be amenable for fabrication at submicrometre length scales and to possess a large and complete bandgap, has recently been introduced by Fan *et al.*³⁷. One embodiment of this type of photonic crystal is shown at the bottom of the left panel of Fig. 6. The crystal is designed to be fabricated in a layered fashion, using two different dielectric materials (for example, Si and SiO₂), with a series of non-intersecting air channels etched at normal incidence through the top surface after growth is completed. (The presence of non-intersecting air channels was actually a constraint imposed on the design and deemed important for easily maintaining high quality during fabrication.) To create a crystal with a larger dielectric contrast (and consequently a larger bandgap), one of the two dielectric materials can be chosen so that it can be removed at the end by selective etching. The sequence of 'growth' steps illustrated in the right panel of Fig. 6 are shown to help the visualization of the basic elements that make up the crystal structure. First, a layer of Si is deposited on a substrate of choice and grooves are etched into the Si layers as shown. The grooves run normal to the page and are arranged in a periodic array. The grooves are then filled with SiO₂ and another Si layer is grown on the previous layer. Long grooves are again etched (which now extend into the first layer) and again back-filled with SiO₂. Repetition of this procedure generates the basic bulk structure of the photonic crystal. The final step is to etch an array of air channels into the top surface of the structure, at normal incidence.

The design of this structure has many degrees of freedom which can be used to optimize the size of the bandgap. Using a dielectric constant of 12.096 for Si at 1.5 μm and 2.084 for silica at the same

frequency gives an optimized complete photonic bandgap of about 14%. A very significant improvement can be made by simply removing the oxide. The resulting photonic-crystal structure is then predicted to possess a sizeable 23% gap³⁷. Moreover, calculations of the predicted transmission spectra for this structure reveal that with a photonic-crystal thickness of only seven layers one can achieve four orders of magnitude reduction in transmission through the bandgap (J. Weitz, S.F., P.R.V. and J.D.J., unpublished results). This structure is currently being fabricated at the 4.5 μm length scale by Kolodziejski, Reif and co-workers at MIT, using optical lithography. Two layers of the photonic crystal have already been successfully grown.

A very new and exciting approach to the design and fabrication of submicrometre 3D photonic crystals involves the creation of a periodic lattice of isolated metallic regions within a dielectric host³⁸. Such 3D metallodielectric photonic crystals have been studied theoretically^{39–42} and can be shown⁴² to have enormous omnidirectional bandgaps approaching 80%. As the metallic regions are not connected and light tends to be attracted away from the metal towards the dielectric, it is hoped that the effects of the losses expected at optical frequencies may not be too important. Efforts are currently underway at the MIT-Lincoln Laboratory to fabricate and measure the losses and other properties of these metallodielectrics.

The challenge

Photonic crystals offer the possibility of controlling and manipulating light by opening a gap in the density of electromagnetic states within a given range of frequencies. Whereas perfect crystals may be valuable for the fabrication of devices such as high-efficiency light-emitting diodes and 3D mirrors, the introduction of a defect in the crystal, either locally or in an extended region, will allow us to generate electromagnetic states with specific properties. The ability to custom-design a state may prove to be essential in the fabrication of laser sources in frequency ranges yet unseen, and novel optical devices such as switches, modulators, filters and interconnects.

From its beginning less than ten years ago, research in the field of photonic crystals has been advancing at high speeds. To maintain the same rate of progress in the next ten years, it will be necessary for experimentalists to overcome the challenges associated with the fabrication of small intricate three-dimensional periodic structures with feature sizes of less than one μm . Only then will photonic crystals be able to fulfill our expectations. □

J. D. Joannopoulos, P. R. Villeneuve and Shanhui Fan are at the Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

1. Joannopoulos, J., Meade, R. & Winn, J. *Photonic Crystals* (Princeton Press, Princeton, NJ, 1995).
2. Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
3. John, S. Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).
4. John, S. Electromagnetic absorption in a disordered medium near a photon mobility edge. *Phys. Rev. Lett.* **53**, 2169–2172 (1984).
5. Anderson, P. W. Absence of diffusion in certain random lattices. *Phys. Rev.* **109**, 1492–1505 (1958).

6. Drake, M. & Genack, A. Observation of nonclassical optical diffusion. *Phys. Rev. Lett.* **63**, 259–262 (1989).
7. Genack, A. & Garcia, N. Observation of photon localization in a three-dimensional disordered system. *Phys. Rev. Lett.* **66**, 2064–2067 (1991).
8. Robertson, W. *et al.* Measurement of photonic band structure in a two-dimensional periodic dielectric array. *Phys. Rev. Lett.* **68**, 2023–2026 (1992).
9. Meade, R., Brommer, K., Rappe, A. & Joannopoulos, J. Nature of the photonic band gap: some insights from a field analysis. *J. Opt. Soc. Am. B* **10**, 328–332 (1993).
10. Meade, R., Brommer, K., Rappe, A. & Joannopoulos, J. Existence of a photonic band gap in two dimensions. *Appl. Phys. Lett.* **61**, 495–497 (1992).
11. Villeneuve, P. & Piché, M. Photonic band gaps in two-dimensional square and hexagonal structures. *Phys. Rev. B* **46**, 4969–4972 (1992).
12. Gruning, U., Lehmann, V., Ottow, S. & Busch, K. Macroporous silicon with a complete two-dimensional photonic band gap centered at 5 μm . *Appl. Phys. Lett.* **68**, 747–749 (1996).
13. Krauss, T., De La Rue, R. & Band, S. Two-dimensional photonic bandgap structures operating at near-infrared wavelengths. *Nature* **383**, 699–702 (1996).
14. Cassagne, D., Jouanin, C. & Bertho, D. Hexagonal photonic-band-gap structures. *Phys. Rev. B* **53**, 7134–7142 (1996).
15. Mekis, A. High transmission through sharp bends in photonic crystal waveguides. *Phys. Rev. Lett.* **77**, 3787–3790 (1996).
16. Slusher, R. Semiconductor microlasers and their applications. *Opt. Photonics News* **4**(2), 8–17 (1993).
17. Meade, R., Brommer, K., Rappe, A. & Joannopoulos, J. Photonic bound states in periodic dielectric materials. *Phys. Rev. B* **44**, 13772–13774 (1991).
18. Yablonovitch, E. Donor and acceptor modes in photonic band structure. *Phys. Rev. Lett.* **67**, 3380–3383 (1991).
19. McCall, S., Platzman, P., Dalichaouch, R., Smith, D. & Schultz, S. Microwave propagation in two-dimensional dielectric lattices. *Phys. Rev. Lett.* **67**, 2017–2020 (1991).
20. Leung, K. Defect modes in photonic band structures: a Green's function approach using vector Wannier functions. *J. Opt. Soc. Am. B* **10**, 303–306 (1993).
21. Maradudin, A. & McGurn, A. in *Photonic Band Gaps and Localization* (ed. Soukoulis, C.) 247–268 (Plenum, New York, 1993).
22. Fan, S. *et al.* Guided and defect modes in periodic dielectric waveguides. *J. Opt. Soc. Am. B* **12**, 1267–1272 (1995).
23. Sigalas, M., Soukoulis, C., Chan, C. & Ho, K. in *Photonic Band Gap Materials* (ed. Soukoulis, C.) 173–202 (Kluwer, Dordrecht, 1996).
24. Birks, T., Atkin, D., Wylangowski, G., Russel, P. & Roberts, P. *Photonic Band Gap Materials* (ed. Soukoulis, C.) 437–444 (Kluwer, Dordrecht, 1996).
25. Villeneuve, P., Fan, S. & Joannopoulos, J. Microcavities in photonic crystals: mode symmetry, tunability, and coupling efficiency. *Phys. Rev. B* **54**, 7837–7842 (1996).
26. Meade, R. *et al.* Novel applications of photonic band gap materials: low loss bends and high Q cavities. *J. Appl. Phys.* **75**, 4753–4755 (1994).
27. Villeneuve, P. *et al.* Air-bridge microcavities. *Appl. Phys. Lett.* **67**, 167–169 (1995).
28. Chen, J., Haus, H., Fan, S., Villeneuve, P. & Joannopoulos, J. Optical filters from photonic band gap air bridges. *IEEE J. Lightwave Tech.* **14**, 2575–2580 (1996).
29. Joannopoulos, J. The almost-magical world of photonic crystals. *Braz. J. Phys.* **26**, 58–67 (1996).
30. Yablonovitch, E., Gmitter, T. & Leung, K. Photonic band structure: the face-centered-cubic case employing nonspherical atoms. *Phys. Rev. Lett.* **67**, 2295–2298 (1991).
31. Chan, C., Ho, K. & Soukoulis, C. Photonic band-gaps in experimentally realizable periodic structures. *Europhys. Lett.* **16**, 563–568 (1991).
32. Sözüer, H. & Haus, J. Photonic bands: simple-cubic lattice. *J. Opt. Soc. Am. B* **10**, 296–302 (1993).
33. Ho, K., Chan, C., Soukoulis, C., Biswas, R. & Sigalas, M. Photonic band gaps in three dimensions: new layer-by-layer periodic structures. *Solid State Commun.* **89**, 413–416 (1994).
34. Sözüer, H. & Dowling, J. Photonic band calculations for woodpile structures. *J. Mod. Opt.* **41**, 231–239 (1994).
35. Özbay, E. *et al.* Micromachined millimeter-wave photonic band-gap crystals. *Appl. Phys. Lett.* **64**, 2059–2061 (1994).
36. Cheng, C. & Scherer, A. Fabrication of photonic band-gap crystals. *J. Vac. Sci. Technol. B* **13**, 2696–2700 (1995).
37. Fan, S., Villeneuve, P., Meade, R. & Joannopoulos, J. Design of three-dimensional photonic crystals at submicron lengthscales. *Appl. Phys. Lett.* **65**, 1466–1468 (1994).
38. Brown, R. & McMahon, O. Large electromagnetic stop bands in metallodielectric photonic crystals. *Appl. Phys. Lett.* **67**, 2138–2140 (1995).
39. McGurn, A. & Maradudin, A. Photonic band structures of two- and three-dimensional periodic metal or semiconductor arrays. *Phys. Rev. B* **48**, 17576–17579 (1993).
40. Pendry, J. Photonic band structures. *J. Mod. Opt.* **41**, 209–229 (1994).
41. Sigalas, M., Chan, C., Ho, K. & Soukoulis, C. Metallic photonic band-gap materials. *Phys. Rev. B* **52**, 11744–11751 (1995).
42. Fan, S., Villeneuve, P. & Joannopoulos, J. Large omnidirectional band gaps in metallodielectric photonic crystals. *Phys. Rev. B* **54**, 11245–11251 (1996).

Acknowledgements. The work was supported by the US NSF through the MRSEC at MIT.

Correspondence and requests for materials should be addressed to J.D.J.

Measurement of the Wigner function of an ensemble of helium atoms

Ch. Kurtsiefer, T. Pfau & J. Mlynek

Fakultät für Physik, Universität Konstanz, Postfach 5560, D-78434 Konstanz, Germany

Beams of atoms can exhibit interference and diffraction phenomena just like waves of light. For a coherent beam of helium atoms in a double-slit experiment, measurements of the quantum-mechanical analogue of the classical phase-space distribution function show that the motion of atoms behaves in a strongly non-classical manner.

Over the past few years, techniques for manipulating the motion of atoms using light fields or micro-structured solids have undergone rapid development¹, and they now provide the fundamental tools for the new field of 'atom optics'. Underlying these techniques is the insight of de Broglie that a particle of mass m moving with a velocity v can be viewed as a matter wave with wavelength $\lambda_{\text{dB}} = 2\pi\hbar/(mv)$. Based on this wave nature, atom interferometers have been demonstrated^{2,3}, which are now finding applications in precision spectroscopy and atomic physics⁴.

Most of the atom interferometers realized to date rely on the fact that the quantum-mechanical state of motion that describes the atoms is such that, at some point in the experiment, they are in a superposition state of being at two different locations x in space, which may be micrometres to centimetres apart. The matter wavefunction corresponding to such a state cannot, however, be measured for a single atom, and the measurement has to be performed on an ensemble of similarly prepared atoms. The full quantum-mechanical description of an ensemble system needs to capture both intrinsic quantum-mechanical uncertainties and possible uncertainties due to a 'classical' lack of knowledge about the initial exact quantum state of each particle in the ensemble. The most complete descriptions are provided by the Wigner function $W(x, p)$ (where p is momentum) which was introduced as early as 1932 (ref. 5) or alternatively the density matrix $\hat{\rho}(x, x')$. These two quantities are related by

$$W(x, p) = \frac{1}{\pi\hbar} \int \langle x + x' | \hat{\rho} | x - x' \rangle e^{i2px'/\hbar} dx'$$

The Wigner function offers a convenient interpretation of the ensemble in the form of a quasi-probability distribution in the phase space defined by position x and momentum p . For quasi-classical states of a system, $W(x, p)$ is identical to the phase-space density of the corresponding classical system⁶. There are, however, states of systems which do not have a classical equivalent at all, such as that of a particle in an atom interferometer when in a linear superposition of states at two spatially separated locations. These states lead to negative values in the quasi-probability distribution $W(x, p)$; in fact, most non-classical states of systems (except those whose variables are described by a gaussian probability distribution) should show negative values in $W(x, p)$. A review of Wigner functions in quantum mechanics can be found in, for example, ref. 7.

It was pointed out by Bertrand and Bertrand⁸ and independently by Vogel and Risken⁹ that the Wigner function of a system can be reconstructed by tomographical methods from the measured distributions of observables that are certain combinations of the non-commuting variables forming the phase space. Additionally,

more sophisticated methods have been developed for the reconstruction of the density matrix as an equivalent complete description of a quantum state^{10–12}.

There have been several experimental measurements of Wigner functions performed for non-classical states of light^{13,14}, as well as measurements of the equivalent function that describes molecular vibrational states¹⁵. Compared to quantum-mechanical states of the electromagnetic field, it is relatively easy to generate a quantum state for a massive particle, such that the wavefunction is a coherent superposition of being at one of two places in space. Very recently, there have been investigations of the state of motion of ions in a harmonic potential by the group of D. Wineland at NIST in Boulder¹⁶, where negativities in the Wigner functions corresponding to Fock states were observed¹⁷. We investigate here the conceptually simple system of a two-peaked matter wave state prepared by coherent illumination of a double slit, where the state of the particle after preparation is governed by free evolution through the apparatus. When the spatially separated wave packets corresponding to each slit show sufficient mutual coherence, the Wigner function shows negative parts. A detailed theoretical analysis of such a system is given by Janicke and Wilkens¹⁸.

Theoretical background

Phase-space tomography requires a set of measured projections of the Wigner function $W(x, p)$ along different directions in phase space. To avoid the problem of mixing the different physical properties of the conjugated variables x and p for the projection along oblique phase-space directions, we introduce a scaling length x_0 and momentum $p_0 = \hbar/x_0$. A rotated basis ($x' = \bar{x} \cos \Theta + \bar{p} \sin \Theta$; $p' = -\bar{x} \sin \Theta + \bar{p} \cos \Theta$) for the phase space can then be expressed by a reduced position $\bar{x} = x/x_0$ and momentum $\bar{p} = p/p_0$, and a mixing angle Θ . The projections or marginal distributions of a motional state described by the Wigner function $W_0(\bar{x}, \bar{p})$ are then given by

$$P_\Theta(x') = \int dp' W_0(\bar{x}, \bar{p})$$

After a motional state $W_0(\bar{x}, \bar{p})$ is prepared, a physical mixing mechanism between x and p has to be introduced to measure the marginal distributions in a set of observables $x'(\Theta)$. This mixing mechanism can be provided by a combination of lenses and free-space propagation for arbitrary complex-valued fields^{19–21}; the key idea is to change a momentum distribution (or a distribution in a mixed base) into a position distribution $P(x)$, which then can be observed with an atom detector. The simplest mixing mechanism is the free evolution of the particle, using the kinetic-energy hamiltonian $H = p^2/2m$. Under this hamiltonian, the Wigner function W_0

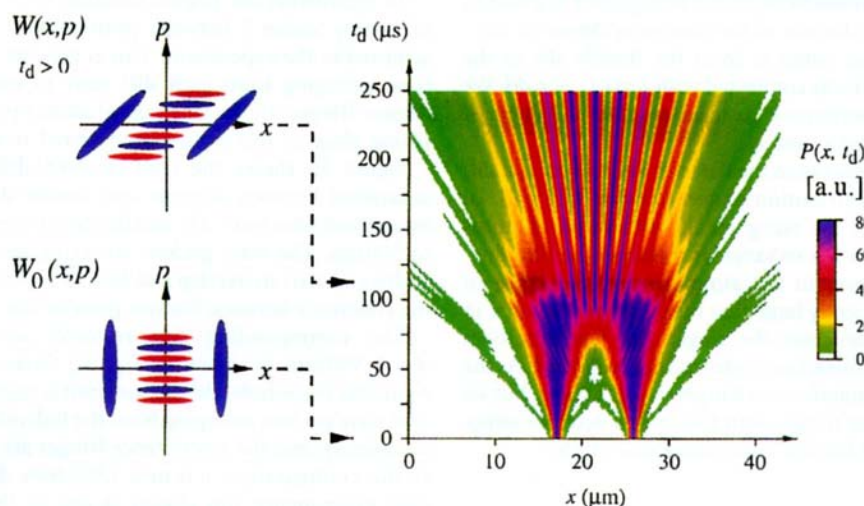


Figure 1 Evolution of a quantum state W_0 initially in a superposition of an atom being behind one of two slits. The two large blue vertical ellipses illustrate the position-momentum (x - p) distribution of atoms isolated behind one or other of the slits (the quasi-classical states); the alternating blue and red horizontal ellipses illustrate the position-momentum distribution of the atoms in the non-classical superposition of states (red corresponds to negative values in the Wigner function)—it is these superposition states that give rise to interference.

After an evolution time t_d , the Wigner function becomes sheared in phase space, and in the corresponding position distribution (main figure) $P(x, t_d)$, interference fringes appear. The plot was obtained by numerical evaluation using Huygens-Fresnel diffraction, where the double slit is illuminated from an incoherent matter wave source with a finite spatial extent of $5\text{ }\mu\text{m}$, corresponding to our experimental situation (a.u., arbitrary units).

of the initial state is sheared in phase space (Fig. 1). This can be easily understood from the classical equation of motion $\partial x/\partial t = p/m$ for a point in phase space. The position distributions from the sheared Wigner function for a set of evolution times t_d may contain enough information to approximately reconstruct the original Wigner function²⁰, as is the case for a set of marginal distributions $P_{\theta}(x')$. In fact, a huge class of hamiltonians can be used to modify a given state W_0 such that its initially hidden aspects can be revealed by only measuring its position distribution after a set of evolution times^{22,23}. Free evolution seems to be just the simplest choice in the case of a massive particle.

Experimental realization

The experimental apparatus is shown in Fig. 2. A beam of metastable helium atoms is generated in a gas-discharge source, which is used in a pulsed operation mode by switching the discharge current. A fraction of the atoms in the beam are in the metastable 2^3S_1 state, produced by electron impact during the switch-on time of $15\text{ }\mu\text{s}$. The longitudinal velocity distribution of the metastable atom beam emerging from the source consists of a fraction of very fast atoms with velocity $v \approx 33,000\text{ m s}^{-1}$ and a fraction at velocities between $1,000$ and $3,000\text{ m s}^{-1}$ corresponding to the thermal energy of the helium gas expanding in a nozzle. The beam passes a skimmer and is collimated by a $5\text{-}\mu\text{m}$ -wide slit. Further downstream, the beam is sent through a micro-fabricated double-slit structure with a slit separation of $8\text{ }\mu\text{m}$ and an opening of $1\text{ }\mu\text{m}$. The experiments are conducted under high vacuum conditions (10^{-6} mbar). The combination of entrance slit and double slit acts as a preparation tool for the transverse motion of the atoms. The atoms then propagate for a distance d to a time- and space-resolving detector²⁴; this detector is based on the fact that metastable (that is, electronically excited) atoms release secondary electrons when striking a metal surface²⁵. Our detector allows us to measure the transverse atomic distribution of metastable atoms with a spatial resolution down to 500 nm , and the arrival time t_f (after leaving the source) of the atoms with an accuracy of 100 ns . From an atom-optical point of view, this corresponds to measurements of atomic distributions corresponding to different de Broglie wavelengths $\lambda_{\text{dB}} = 2\pi\hbar/(mv)$, where m is

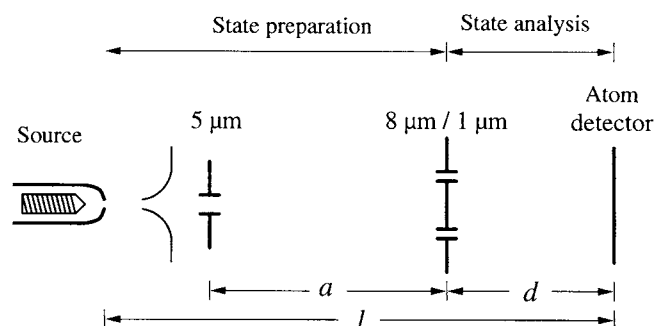


Figure 2 Diagram of apparatus used to observe atomic interference patterns. The single slit is $5\text{ }\mu\text{m}$ wide; the double slit has a slit separation of $8\text{ }\mu\text{m}$ and a slit width of $1\text{ }\mu\text{m}$. The experiment is conducted under high vacuum conditions. See text for details.

the mass of a He atom and $v = l/t_f$ (l is the source-detector distance). The mean number of detected atoms per source pulse varies from 0.01 to 0.2 ; the source is repeatedly fired, and the obtained distribution summed up.

For the very fast atoms, the wave packet associated with the transverse motion has not enough time to spread out in space, so the transverse distribution of the atom stays as prepared by the double-slit structure. In the language of atom optics, this may be regarded as a geometrical shadow of the structure on the detector. Because the double slit is illuminated from a single slit rather than a source emitting a plane matter wave, the observed shadow is magnified by a factor $(a + d)/a$, where a is the distance between the single slit and the double slit. In the experiment, we use this information to gauge the length scaling of the detector.

Time-resolved diffraction patterns

Because the longitudinal motion of the atoms at velocities v of several thousand metres per second may be treated completely

classically, the apparatus shown in Fig. 2 can be regarded as a tool to investigate the spatial distribution of the transverse degree of freedom for a set of evolution times t_d from the double slit to the detector, which is geometrically connected with t_f via $t_d = t_f d/l$. We therefore describe the interference patterns that we obtained in terms of a free evolution for a time t_d .

For a set of separations d between the detector and the double slit, we obtained the atomic distribution patterns shown in Fig. 3. In Fig. 3a ($d = 148$ mm), one can recognize the diffraction of atoms from the individual slits with a rectangular transmission function. For the velocity distribution in the atomic beam, the range of evolution times t_d is sufficiently large that the detector is located in the far field for diffraction from the single slits; the side lobes corresponding to the first diffraction order can be recognized. As the two wave packets overlap, interference fringes should appear, but we were not able to resolve the fringes with the chosen detector setup. The total integration time for this experiment was 14.5 h.

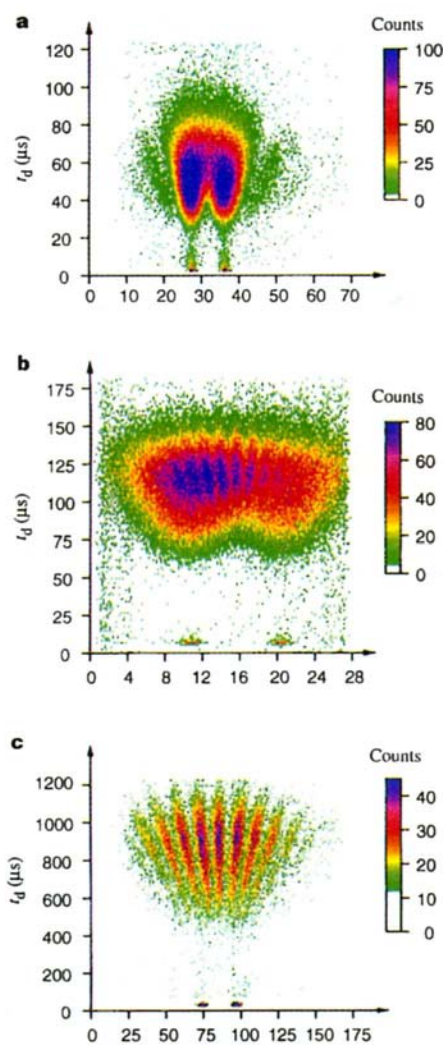


Figure 3 Distribution of atoms in the detection plane as a function of their velocities. This corresponds to different evolution times for a partial coherent matter wave packet after the double slit. **a**, For a separation $d = 148$ mm between detector and double slit (see Fig. 2). The diffraction pattern is determined by the diffraction from the single slits. **b**, For the transition regime at $d = 248$ mm. Partial waves from both slits start to interfere. **c**, For a separation $d = 1,950$ mm (far field). The interference orders are clearly separated, and their separation grows linearly with the evolution time.

To reconstruct the Wigner function W_0 , a sufficiently large range of mixing angles θ between position and momentum has to be achieved in the experiment. This is the case when the two wavelets (one emerging from each slit) start to overlap and form interference fringes. In the language of atom optics, interference effects taking place in this regime are referred to as Fresnel diffraction.

Figure 3b shows the time-resolved diffraction pattern for a separation between detector and double slit of $d = 248$ mm; the integration time was 7.2 h, and the detector resolution was increased to 500 nm. The wave packets emerging from each slit increase in width and start to overlap and form interference fringes, indicating the coherence between the two parts of the wavefunction.

The corresponding time-resolved atomic distribution for $d = 1,950$ mm is shown in Fig. 3c. Given the slit separation of 8 μm , the Fraunhofer diffraction limit is reached for the slow atoms. The wave packets emerging from the individual slits overlap almost completely, and the interference fringes are most clearly observed. In this configuration, a is now 1,052 mm, differing from the near-field experiments; this change is due to the requirements of the apparatus. To compensate for the small atom flux seen by the detector because of its large separation from the source, we had to extend the integration time in this experiment to 49 h.

Reconstruction of the Wigner function

The transformation algorithm used to reconstruct the prepared Wigner function is similar to that described in ref. 18. Free evolution of the particle for a time t_d leads to a corresponding position distribution $P(\tilde{x}, t_d)$, which is connected with a rotated marginal

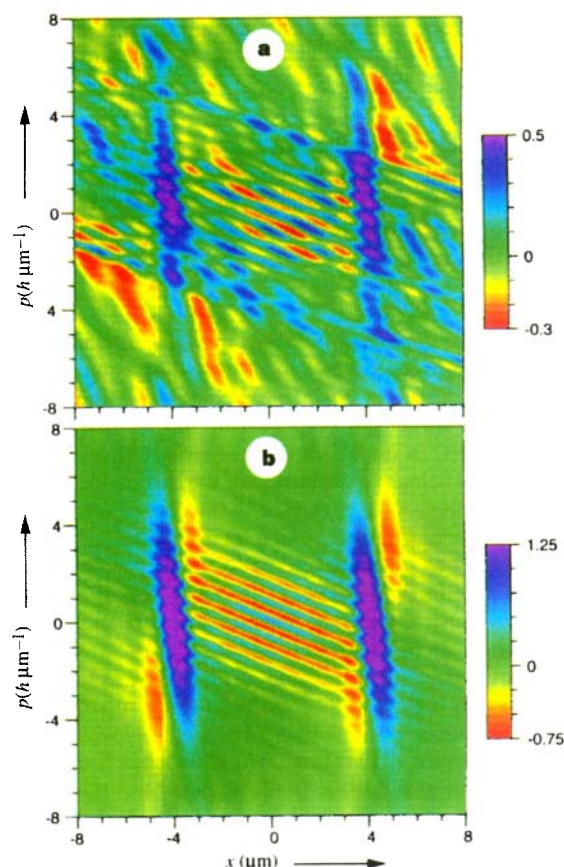


Figure 4 a, Reconstructed Wigner function $W(x,p)$ of the investigated atomic ensemble, obtained from the experimental distribution shown in Fig. 3b using the inverse Radon transformation^{13,18}. **b**, Reconstruction of $W(x,p)$ from the numerical distribution in Fig. 1.

distribution $P_{\Theta}(\bar{x})$ by:

$$P_{\Theta}(\bar{x}) = P\left(\frac{\bar{x}}{\cos \Theta}, \frac{m\bar{x}_0^2}{\hbar} \tan \Theta\right)$$

As there is only access to positive evolution times t_d , free evolution only gives access to rotated marginal distributions P_{Θ} for angles Θ between 0 and $\pi/2$. Theoretical investigations^{18,19} have shown that the use of additional optical elements (lenses) could allow the complete set of mixing angles Θ ($-\pi/2$ to $\pi/2$) to be investigated; this is necessary to reconstruct a completely unknown state. Alternatively, additional symmetry assumptions can be made for the Wigner function investigated, such as left/right symmetry or time-reversal symmetry. Here we use assumption of symmetry in position space for the reconstructed Wigner function.

Furthermore, only a limited range of evolution times t_d is available in an experiment. But by choosing an appropriate range of evolution times, the interesting information of a system should be contained in the observed distributions. Here, the information is the coherence property of two spatially separated wave packets.

Another problem for the reconstruction process is the coupling between longitudinal and transverse motion, as we illuminate the double slit from a single slit instead of a pure plane-wave source. This effect can be taken into account by rescaling the transverse position in the detection plane by the factor $a/(a+d)$ of geometrical magnification; the evolution-time-dependent mixing process between position and momentum is preserved.

The reconstructed Wigner function we obtained from the diffraction data shown in Fig. 3b is shown in Fig. 4a, where we chose a scaling length $x_0 = 2 \mu\text{m}$ and a cut-off frequency $r_c = 10$ (see ref. 18 for details). The two positive regions, corresponding to the spatial distribution of the atoms during preparation, may be seen. The separation between the regions ($\sim 8 \mu\text{m}$) corresponds to the position distribution in the plane of the double slit; the information relevant to determining this feature of the Wigner function is mainly contributed by the very fast atoms. The coherence between the two spatially separated parts of the Schrödinger field at the position of the double slit then leads to interference, which is manifested in oscillations in the Wigner function in the region between the two main 'blobs'. In this region, the reconstructed Wigner function takes negative values, indicating a property which cannot be obtained by classical phase-space distributions and revealing the quantum nature of the observed ensemble of atoms.

But, in contrast to what is expected¹⁸, the reconstructed Wigner function appears sheared, and regions of negative value appear close to the two large positive blobs. This is consistent with the reconstruction shown in Fig. 4b which was obtained from the numerically obtained distributions shown in Fig. 1, where the largest evolution time was restricted to 175 μs ; the shear and the regions of negative value close to the positive blobs are therefore a consequence of the reconstruction algorithm. The non-physical shear disappears if longer evolution times t_d for the observed marginal distributions $P(x, t_d)$ are allowed, as additional numerical tests have shown.

More sophisticated reconstruction techniques than the inverse Radon transform^{13,18} could therefore be usefully applied to our data. Promising tools for reconstructing the Wigner function from such an incomplete set of measured marginal distributions may be a technique using bayesian analysis (described in ref. 11 for application in optical homodyne tomography) or maximum entropy algorithms (as suggested by V. Bužek *et al.*¹²).

Here we have reported on the measurement of the Wigner function of a massive particle wave packet formed by partially coherent illumination of a double slit; we have been able to identify negative parts in the oscillatory part of the Wigner function corresponding to a superposition of macroscopically separated parts of the matter wave field. In quantum optics, such a state is sometimes referred to as a 'Schrödinger cat state'¹⁶.

Our technique provides a convenient approach to detection of particles for which the key observables, position and momentum, can be mixed controllably. Such a choice of a detection base would be necessary for Bell-type experiments²⁶ with entangled atom-photon pairs, which can be efficiently produced and detected²⁷. The measurement technique reported here could be used for measuring the atomic part of these pairs. It may also be useful in characterizing the motional state of atoms emerging from one of the new atomic sources currently under discussion in the context of Bose condensates. $\square$

Received 4 November 1996; accepted 21 January 1997.

- Adams, C. S., Sigel, M. & Mlynek, J. *Phys. Rep.* **240**, 143–210 (1994).
- Carnal, O. & Mlynek, J. *Phys. Rev. Lett.* **66**, 2689–2692 (1991).
- Keith, D. W., Ekstrom, C. R., Turchette, Q. A. & Pritchard, D. E. *Phys. Rev. Lett.* **66**, 2693–2696 (1991).
- Berman, P. (ed.) *Atom Interferometry* (Academic, San Diego, 1997).
- Wigner, E. *Phys. Rev.* **40**, 749–759 (1932).
- Royer, A. *Foundat. Phys.* **19**, 3–32 (1989).
- Hillery, M., O'Connell, R. F., Scully, M. O. & Wigner, E. P. *Phys. Rep.* **106**, 121–167 (1984).
- Bertrand, J. & Bertrand, P. *Foundat. Phys.* **17**, 397–405 (1987).
- Vogel, K. & Risken, H. *Phys. Rev. A* **40**, 2847–2849 (1989).
- D'Ariano, G. M., Macciavello, C. & Paris, M. G. A. *Phys. Rev. A* **50**, 4298–4302 (1994).
- Tan, S. M. *J. Mod. Opt.* (submitted).
- Busžek, V., Adam, G. & Drobný, G. *Ann. Phys. (NY)* **254**, 37–97 (1996).
- Smithey, D. T., Beck, M., Raymer, M. G. & Faridani, A. *Phys. Rev. Lett.* **70**, 1244–1247 (1993).
- Breitenbach, G. *et al. J. Opt. Soc. Am. B* **12**, 2304–2309 (1995).
- Dunn, T. J., Walmsley, I. A. & Mukamel, S. *Phys. Rev. Lett.* **74**, 884–887 (1995).
- Monroe, C., Meekhof, D. M., King, B. E. & Wineland, D. J. *Science* **272**, 1131–1135 (1996).
- Leibfried, D., King, B. E., Monroe, C., Itano, M. & Wineland, D. J. *Phys. Rev. Lett.* **77**, 4281–4285 (1996).
- Janicke, U. & Wilkens, M. *J. Mod. Opt.* **42**, 2183–2199 (1995).
- Raymer, M. G., Beck, M. & McAlister, D. F. *Phys. Rev. Lett.* **72**, 1137–1140 (1994).
- Lohmann, A. W. *J. Opt. Soc. Am. A* **10**, 2181–2186 (1993).
- Kienle, S. H., Freyberger, M., Schleich, W. P. & Raymer, M. G. in *Experimental Metaphysics: Quantum Mechanical Studies for Abner Shimony* (eds Cohen, R. S., Horne, M. A. & Stachel, J.) 121–133 (Kluwer, Dordrecht, 1997).
- Richter, Th. & Wünsche, A. *Phys. Rev. A* **53**, 1974–1977 (1996).
- Leonhard, U. & Raymer, M. G. *Phys. Rev. Lett.* **76**, 1985–1989 (1996).
- Kurtsiefer, Ch. & Mlynek, J. *Appl. Phys. B* **64**, 85–90 (1996).
- Hotop, H. in *Methods of Experimental Physics* Vol. 29B (eds Dunning, F. B. & Hulet, R. G.) 191–215 (Academic, San Diego, 1995).
- Pfau, T., Kurtsiefer, Ch. & Mlynek, J. *Quant. Semicond. Opt.* **8**, 665–671 (1996).
- Kurtsiefer, Ch. *et al. Phys. Rev. A* (submitted).

Acknowledgements. We thank U. Janicke and M. Wilkens for discussions. This work was supported by the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to Ch.K. (e-mail: christian.kurtsiefer@uni-konstanz.de).

Origin of asteroid rotation rates in catastrophic impacts

Stanley G. Love & Thomas J. Ahrens

Seismological Laboratory 252-21, California Institute of Technology, Pasadena, California 91125, USA

The rotation rates of asteroids, which are deduced from periodic fluctuations in their brightnesses¹, are controlled by mutual collisions^{2–8}. The link between asteroid spin and collision history is usually made with reference to impact experiments on centimetre-scale targets, where material strength governs the impact response^{2,3,9–11}. Recent work, however, indicates that for objects of the size of most observed asteroids (≥ 1 km in diameter), gravity rather than intrinsic strength controls the dynamic response to collisions^{12–14}. Here we explore this idea by modelling the effect of impacts on large gravitating bodies. We find that the fraction of a projectile's angular momentum that is retained by a target asteroid is both lower and more variable than expected from laboratory experiments, with spin evolution being dominated by 'catastrophic' collisions that eject ~ 50 per cent of the target's mass. The remnant of an initially non-rotating silicate asteroid that suffers such a collision rotates at a rate of ~ 2.9 per day, which is close to the observed mean asteroid rotation rate of ~ 2.5 d⁻¹. Moreover, our calculations suggest that the observed trend in the mean spin frequency for different classes of asteroids⁴ (2.2 d⁻¹ for C-type asteroids, 2.5 d⁻¹ for S-type, and 4.0 d⁻¹ for M-type) is due to increasing mean density, rather than increasing material strength.

The computer model¹⁴, a three-dimensional smoothed particle hydrodynamics (SPH) code, neglects intrinsic strength (justifiable for the silicate targets ≥ 10 km in diameter treated here^{12–14}) but includes a rigorous treatment of gravitation. The SPH algorithm¹⁵ models a continuous medium using discrete particles which are mathematically 'smoothed' out into the surrounding volume; unlike grid-based methods, it is good for simulating impacts because it remains accurate even when the collision partners suffer extreme geometrical distortion. Figure 1 illustrates the impact geometry and coordinate system. Throughout each numerical run, the gravitational force and potential are calculated for each particle using the masses and positions of all other particles. We explicitly compute the projectile's contact with the target, compression at the impact site, ejecta launch, and propagation of the impact shock wave to the target antipode on a timescale of seconds to minutes. Later evolution of the system is primarily ballistic and is treated analytically.

In the ballistic phase, we calculate each particle's kinetic energy, $K_i = \frac{1}{2} m_i v_i^2$, where m is the particle's mass and v is its velocity in the centre-of-mass reference frame. We also find the energy of gravitational binding between each particle and all the other particles in the system, $W_i = \sum_{j \neq i} -G m_i m_j / r_{ij}$, where G is the universal gravitational constant and r is the separation between each pair of particles. Particles with $K_i < |W_i|$ do not escape (although some may be launched on suborbital trajectories, reaccreting hours after the collision); these form the final 'rubble pile'. We calculate the position and speed of the rubble pile's centre of mass and find the total angular momentum of all bound particles about that centre assuming no important gravitational torques between the rubble pile and the escaping ejecta. The rubble pile's angular momentum is translated to a rotation rate by assuming that the bound particles reaccrete into a homogeneous sphere with moment of inertia $I = \frac{2}{5} M R^2$, where M is its mass and R its radius. The results are summarized in Table 1 and plotted as a function of final-to-initial target mass ratio (μ) in Fig. 2a. Ejected particles escape individually;

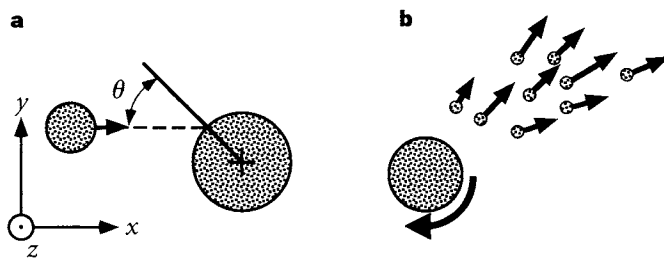


Figure 1 a, Cartoon illustrating the numerical impact simulation geometry. The coordinate system is shown at left; the problem is mirror symmetric about the x - y plane. The impact angle, θ , is shown. Most trials (Table 1) employ $\theta = 45^\circ$ (the most likely value), but two simulations treat grazing (75°) incidence and one treats a near-vertical (15°) impact. The average mutual impact speed between asteroids is ~ 5 km s⁻¹; we employ 5 km s⁻¹ for most trials, but investigate two at 7 km s⁻¹ and one at 3 km s⁻¹. The projectile and target materials are represented by the Tillotson^{20,21} equation of state for granite (density 2,680 kg m⁻³, consistent with that of S-class asteroid 243 Ida²²), except for two simulations using iron (7,800 kg m⁻³) and two with dry tuff (1,700 kg m⁻³). The diameters of the initially non-rotating targets are 10, 31.6, 100, 316 and 1,000 km. The model's resolution is coarse (15 mass elements across the target diameter), but sufficient to resolve the impact shock and generally to map the exchange of energy between the collision partners. **b**, After the impact, some target material escapes in a highly asymmetric ejecta cloud. The remaining gravitationally bound rubble pile retains a net rotation. Because the simulation is mirror symmetric across the x - y plane (**a**), the rubble pile's angular momentum ($\mathbf{L}$) vanishes except about the z axis. It is computed using $\mathbf{L} = \mathbf{L}_z = \sum m(r_x v_y - r_y v_x) \hat{z}$, where the summation includes all mutually bound particles, m is each particle's mass, and r and v are the x and y separation and velocity of each particle relative to the centre of mass.

we cannot define their rotation speeds for comparison with laboratory results on impact ejecta spin rates¹⁰.

The rotation of a collisionally mature asteroid is the result of a 'random walk' of many impacts^{2,4,7}. Applying the present model to observed asteroid rotation rates is uncertain because it neglects initial target spin and because it shows the effect of a single impact (one step of the random walk) rather than the sum of many. Initial spin, however, has only a minor effect on mass ejection⁶. Furthermore, if impacts occur at random orientations, the effects of initial target rotation may average to zero. Combining the curve in Fig. 2a with the observation that target mass removal is proportional to impactor mass for a given target size¹⁴ and including the impactor mass distribution (the relative number N of asteroids with masses in the increment between M and $M + dM$, $N(M, M + dM) \propto M^{-1.7 \pm 0.2} dM$ (refs 5, 9)) allows us to compare the importance of different sized impacts in controlling asteroid spin (Fig. 2b). Small erosive events are frequent but have little effect on the spin. Impacts removing more than ~ 0.75 of the target's mass alter rotation significantly, but not enough to compensate for their rarity. Between those extrema, occasional catastrophic impacts leaving remnants with 0.40–0.65 of their original masses are the most effective in determining asteroid rotation rates and thus represent the 'step size' of the random walk. According to the present model, such impacts produce (on initially non-rotating targets) spin rates of 1.8–4.2 d⁻¹. This range is consistent with the ~ 2.5 d⁻¹ observed asteroidal mean rotation rate^{2,4}. Agreement between these parameters makes sense if catastrophic impacts destroy all 'memory' of initial spin, so that each asteroid exhibits the rotation imparted by its last big collision. Because impact ejecta trajectories in this model are highly directional (mainly down-range), we cannot directly compare the present results with the

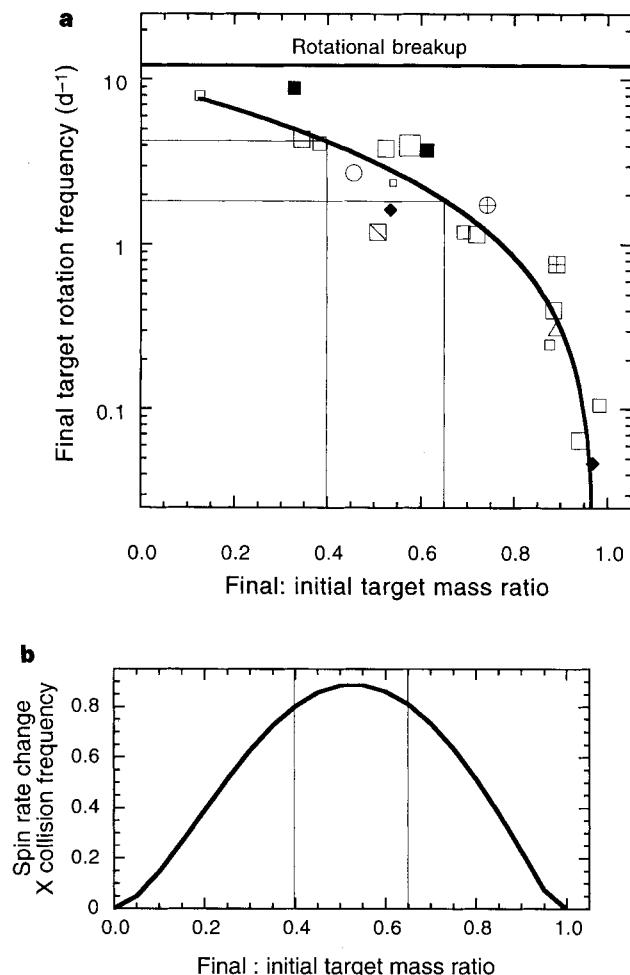


Figure 2 a, Final target rotation frequency f as a function of the ratio, μ , of the target's post-impact mass to its initial mass. The unshaded points denote granite targets. Symbols without lines inside represent 45° impacts, crossed symbols 75°, and the slashed symbol 15°. Square symbols denote 5 km s^{-1} impact speed, round 7 km s^{-1} , and triangular 3 km s^{-1} . The five different symbol sizes represent, from smallest to largest, the 10, 31.6, 100, 316 and 1,000 km target body diameters. The filled squares represent iron and the filled diamonds dry-tuff targets; these trials employed 100-km targets struck at 5 km s^{-1} and 45°. The final rotation rate does not depend strongly on the absolute target size or on the impact velocity. Grazing and near-vertical impacts yield respectively faster and slower rotation for the same mass loss. The heavy horizontal line marks the 11.9 d^{-1} maximum spin rate (corresponding to rotational breakup) for strengthless spherical bodies of density 2,680 kg m^{-3} (see text). The heavy curved line is a fit to the points for 45° impacts on granite. Its formula is $f(\text{in } \text{d}^{-1}) = 9.562 - 15.95\mu + 6.292\mu^2$. **b**, Relative importance (in arbitrary units, with dimensions of time^{-2}) of impacts of different severity in determining asteroid rotation rate. Shown is the product of the change in spin rate imparted to a target by impacts of various sizes (from the present model) and the relative frequency of such collisions, assuming an asteroid mass-frequency distribution with a power-law slope near -1.7 (see text). Collisions that leave target remnants with 0.40–0.65 of their original mass and correspond (according to **a**) to a change in spin frequency of 1.8–4.2 d^{-1} control rotation rate evolution.

'angular momentum drain' de-spinning scheme of Dobrovolskis and Burns⁶ or the 'angular momentum splash' of Cellino *et al.*³, which are based on axisymmetric launch of ejecta from asteroid impact sites. The former mechanism may, however, work to slow asteroid rotation during the intervals between catastrophic collisions.

Figure 2a demonstrates that final spin rate is related to the fraction of retained target mass and the impact angle, but not absolute target size or impact speed in the ranges investigated. Grazing (75°) impacts yield spin rates roughly twice those at 45° for the same degree of mass loss; near-vertical (15°) impacts produce spin frequencies a factor of ~ 2 smaller. The absence of a strong size effect contrasts with the notion that a change in collision physics with size might be responsible for the significant number of slowly rotating asteroids of diameter near 100 km (refs 2, 4). That population is probably not an artefact of the changeover from strength to gravity dominance, as suggested by Binzel *et al.*⁴, given the recent theoretical^{12–14} and observational^{16,17} evidence that the transition occurs at a much smaller size. Instead, we suggest that it arises from a change in the composition or differentiation of asteroids at 100 km diameter, or from a change in the population of projectile asteroids at the size appropriate to drastically alter the spin of 100-km targets.

The line at the top of Fig. 2a indicates an equatorial rotation velocity equal to the low orbit velocity. This is the rotational breakup speed (and maximum rotation frequency) for strengthless gravitating objects. For a sphere, the breakup spin rate (in d^{-1}) is given by $f_{\text{max}} = 86,400 \sqrt{G\rho/3\pi}$, where ρ and G are the density and the gravitational constant in mks units. For increasingly large impacts, final spin rates approach but do not reach the breakup limit. This effect suggests that density may influence spin rate^{7,16}. In

fact, the observed mean spin frequencies of C-, S- and M-class asteroids (2.2 d^{-1} , 2.5 d^{-1} and 4.0 d^{-1} , respectively⁴) fall in nearly the same proportion as the square roots of their presumed densities ($\sim 2,000 \text{ kg m}^{-3}$ for the 'carbonaceous' C class, $\sim 2,700 \text{ kg m}^{-3}$ for the S class, and $\sim 7,800 \text{ kg m}^{-3}$ for the 'metallic' M class). We have investigated the relationship between spin and density in our model. Two simulations using iron ($\rho = 7,800 \text{ kg m}^{-3}$) targets showed spin rates an average of 1.9 times faster than for granite with the same mass loss. This result is consistent (within the scatter of Fig. 2a) with $\rho^{1/2}$ scaling, which predicts a factor of 1.7. Analogously, two runs using dry tuff ($\rho = 1,700 \text{ kg m}^{-3}$) yielded spin rates averaging 0.8 times that of granite, also consistent with $\rho^{1/2}$ scaling, which predicts 0.80. We suggest on the basis of this favourable agreement that the observed trend in rotation rate from C- to S- to M-class asteroids, which has been attributed to increasing impact strength⁴, arises instead from increasing density^{7,16}. We also predict, barring possible effects of lower collision speeds, that collisionally mature Kuiper-belt objects ($\rho \sim 1,000 \text{ kg m}^{-3}$) should have a mean spin rate near 1.5 d^{-1} .

Table 1 includes each rubble pile's angular momentum expressed as a fraction (ζ) of that carried by the projectile before the collision. The ζ parameter is a central pillar of asteroid spin evolution models^{2,6–8}. Experimental values of ζ range from 0.1 to 0.7 (refs 10, 11); Farinella *et al.*² use 0.01–0.3 in their recent work. In the present model ζ ranges from 0.01 to 0.1, varying with impact speed, impact angle, target size, target composition, and fraction of target mass removed. Assuming a constant value of ζ thus appears problematic for models where those parameters are permitted to vary. Note that ζ in large asteroid collisions appears to be significantly smaller than observed in centimetre-scale laboratory impacts.

Table 1 Asteroid collision model outcomes

Target diam. (km)	Projectile diam. (km)	Target remnant		
		μ	Spin rate (d ⁻¹)	ξ
10.0	0.823	0.54	2.38	0.062
31.6	2.60	0.88	0.248	0.046
31.6	4.99	0.12	7.97	0.008
100	8.23	0.98	0.106	0.076
100	15.8	0.69	1.19	0.068
100	19.8	0.38	4.11	0.043
100	22.0	0.61	3.74	0.064
100	26.4	0.33	8.88	0.031
100†	8.23	0.97	0.047	0.033
100†	15.8	0.53	1.62	0.060
316	52.1	0.94	0.065	0.017
316	62.6	0.88	0.402	0.054
316	78.0*	0.89	0.311	0.037
316	78.0†	0.51	1.19	0.091
316	78.0	0.72	1.16	0.058
316	78.0‡	0.89	0.764	0.040
316	78.0§	0.46	2.73	0.045
316	78.0¶	0.74	1.74	0.048
316	93.3	0.52	3.82	0.066
316	100	0.34	4.36	0.030
1,000	464	0.58	4.02	0.066

The nominal case is impact angle 45°, impact velocity 5 km s⁻¹, granite projectile and target; μ is the final-to-initial target mass ratio; the ξ parameter is defined in the text; the impact angle θ is shown in Fig. 1a. * Impact velocity 3 km s⁻¹. † Impact angle 15°. ‡ Impact angle 75°. § Impact velocity 7 km s⁻¹. || Iron projectile and target. ¶ Dry tuff projectile and target.

Further investigation of the rotational evolution of gravity-dominated objects through catastrophic impacts remains to be done. Of greatest interest are the effects of initial target spin and of multiple impacts, the results of improved spatial resolution in computer models, and the influence of catastrophic impacts on the rotation of planet-sized bodies. The last may be important in regard to understanding planetary spins as the results of giant collisions during the late stages of planet accretion¹⁸. □

Received 10 October 1996; accepted 21 January 1997.

- Harris, A. W. & Lupishko, D. F. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 39–53 (Univ. Arizona Press, Tucson, 1989).
- Farinella, P., Davis, D. R., Paolicchi, P., Cellino, A. & Zappalà, V. *Astron. Astrophys.* **253**, 604–614 (1992).
- Cellino, A., Zappalà, V., Davis, D. R., Farinella, P. & Paolicchi, P. *Icarus* **87**, 391–402 (1990).
- Binzel, R. P., Farinella, P., Zappalà, V. & Cellino, A. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 416–441 (Univ. Arizona Press, Tucson, 1989).
- Davis, D. R., Weidenschilling, S. J., Farinella, P., Paolicchi, P. & Binzel, R. P. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 805–826 (Univ. Arizona Press, Tucson, 1989).
- Dobrovolskis, A. R. & Burns, J. A. *Icarus* **57**, 464–476 (1984).
- Harris, A. W. *Icarus* **40**, 142–153 (1979).
- Burns, J. A. & Safronov, V. S. *Mon. Not. R. Astron. Soc.* **165**, 403–411 (1973).
- Fujiwara, A. et al. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 240–265 (Univ. Arizona Press, Tucson, 1989).
- Fujiwara, A. & Tsukamoto, A. *Icarus* **48**, 329–334 (1981).
- Yanagisawa, M., Eluszkiewicz, J. & Ahrens, T. J. *Icarus* **94**, 272–282 (1991).
- Holsapple, K. A. *Planet. Space Sci.* **42**, 1067–1078 (1994).
- Ryan, E. V. & Melosh, H. J. *Eos* **76**, F336 (1995).
- Love, S. G. & Ahrens, T. J. *Icarus* **124**, 141–155 (1996).
- Monaghan, J. J. *Annu. Rev. Astron. Astrophys.* **30**, 543–574 (1992).
- Harris, A. W. *Lunar Planet. Sci.* **XXVII**, 493–494 (1996).
- Botke, W. F. Jr & Melosh, H. J. *Nature* **381**, 51–53 (1996).
- Dones, L. & Tremaine, S. *Science* **259**, 350–354 (1993).
- Botke, W. F. Jr, Nolan, M. C., Greenberg, R. & Kolvoord, R. A. *Icarus* **107**, 255–268 (1994).
- Tillotson, J. H. *General Atomic Rep. GA-3216* (General Atomic div. General Dynamics corp., San Diego, California, 1962).
- Melosh, H. J. *Impact Cratering: A Geologic Process* (Oxford Univ. Press, New York, 1989).
- Belton, M. J. S. et al. *Nature* **374**, 785–788 (1995).

Acknowledgements. We thank P. Farinella for suggesting this line of study, and A. W. Harris and H. J. Melosh for reviews of this Letter. The Cray supercomputer used in this investigation was provided by funding from the NASA Offices of Mission to Planet Earth, Aeronautics, and Space Science; the research itself was supported by NASA.

Correspondence and requests for materials should be addressed to S.G.L. (e-mail: sglove@gps.caltech.edu).

Colossal magnetoresistance in Cr-based chalcogenide spinels

A. P. Ramirez*, R. J. Cava*† & J. Krajewski*

* Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

† Chemistry Department and Princeton Materials Institute, Bowen Hall, Princeton University, Princeton, New Jersey 08540, USA

Manganese oxides with a perovskite structure¹ exhibit a transition between a paramagnetic insulating phase and a ferromagnetic metal phase. Associated with this transition is an effect known as colossal magnetoresistance^{2–5} (CMR)—in the vicinity of the transition temperature, the materials exhibit a large change in resistance in response to an applied magnetic field. Such an effect, if optimized, might find potential application in magnetic devices. But the criteria for achieving (and hence optimizing) CMR are not clear, presenting a challenge for materials scientists. The accepted description of CMR in the manganite perovskites invokes the ‘double-exchange’ mechanism, whereby charge transport is enhanced by the magnetic alignment of neighbouring Mn ions of different valence configuration (Mn³⁺ and Mn⁴⁺), and inhibited by the formation of charge-induced localized lattice distortions^{6,7}. Here we report the existence of a large magnetoresistive effect in a class of materials—Cr-based chalcogenide spinels—that do not possess heterovalency, distortion-inducing ions, manganese, oxygen or a perovskite structure. The realization of CMR in compounds having a spinel structure should open up a vast range of materials for the further exploration and exploitation of this effect.

The metal–insulator transition long known to exist¹ in manganite perovskites with the compound formula A_{1–x}B_xMnO₃, where A is a rare-earth element and B a divalent cation, is unusual in that the low-temperature state is metallic and ferromagnetic whereas the high-temperature state is insulating and weakly paramagnetic. This gives rise to a large negative magnetoresistance (MR) near the Curie temperature, T_C. It has only recently been recognized^{2–5} that this effect might offer advantages over the more well-developed giant magnetoresistance (GMR)⁸ for applications of reading magnetic media. The phenomenon of MR arising from large resistance drops associated with critical ferromagnetic fluctuations is now broadly referred to as colossal magnetoresistance (CMR). Although GMR exceeds CMR in the manganite perovskites at room temperature and low fields, CMR does not saturate until applied fields at least of the order of 6 T where the MR, when normalized to the high field resistance, can exceed that of GMR by orders of magnitude. It has been suggested that the field scale over which CMR materials can be applied will be pushed to lower values by field-amplification⁹ or spin-valving techniques^{10,11}. Given the interest in the CMR phenomenon, it is important to know both from a materials-engineering as well as a basic-physics standpoint, to what extent the effect occurs in other compounds. We investigate here a spinel-structure chalcogenide compound with composition AB₂X₄ where A and B are transition-metal cations occupying tetrahedral and octahedral sites, respectively, and X is a chalcogen. This class of compounds differs from the perovskite oxides in several ways. First, there are many ferromagnetic members of this family¹². Second, there is the possibility of varying both anion as well as cation species while maintaining the spinel structure. Third, unlike in the perovskites where the A-cation acts mainly to donate charge and to control the unit cell volume, in the spinels it is common to have A-cation-derived states near the Fermi level. Last, the electronic states responsible for conduction are not related to Mn³⁺–Mn⁴⁺ mixtures as in the manganite perovskites.

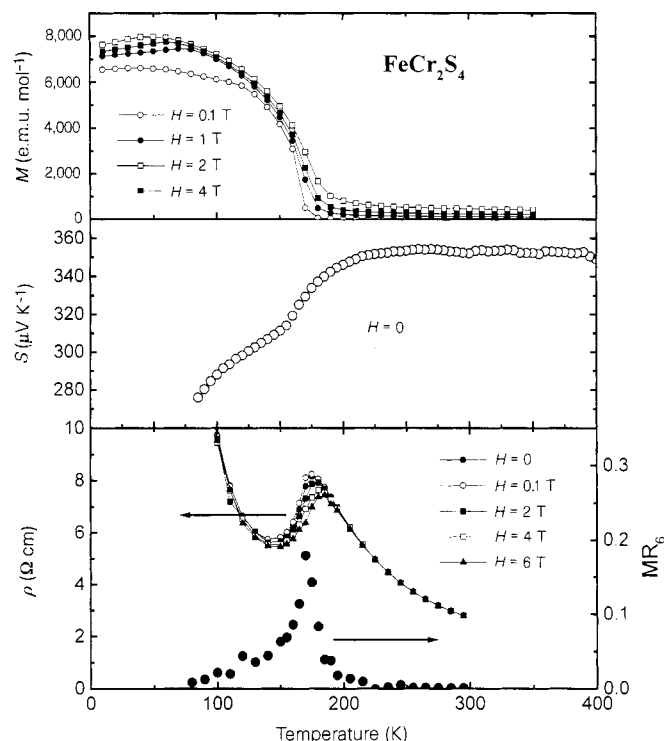


Figure 1 Magnetization (M), thermopower (S), resistivity (ρ) and magnetoresistance ($MR_H = (\rho(0) - \rho(H))/\rho(0)$), versus temperature for FeCr_2S_4 .

Here we extend transport and magnetization data to high fields for the solid solution compounds $\text{Fe}_{1-x}\text{Cu}_x\text{Cr}_2\text{S}_4$ for $x = 0$ and 0.5 . Although these particular compounds have been well studied^{13–20}, there is, to our knowledge, little data extending to fields much larger than 1 T. Both endmembers, $x = 0, 1$, have a normal spinel cation distribution where Cr ions occupy the octahedral sites and Fe and Cu occupy tetrahedral sites. The magnetism and electronic transport have been characterized over the whole range of x . At $x = 0$, FeCr_2S_4 is semiconducting and ferrimagnetic with a Curie temperature, $T_c \approx 180$ K. At $x = 1$, CuCr_2S_4 is metallic and ferromagnetic (FM) with $T_c \approx 400$ K. One of the more striking aspects of this system is the x -dependence of the thermopower, S , which, at room temperature, crosses zero at $x \approx 0.2$ and 0.5 and is very large at $x = 0$ ($S \approx +350 \mu\text{V K}^{-1}$) and for x just below and above 0.5 (-250 and $+300 \mu\text{V K}^{-1}$, respectively)⁶. This has been interpreted to arise from depopulation of the Fe^{2+} levels above s - p valence bands. The Cu ion is thought to be monovalent and hence nonmagnetic in this system.

In Figs 1 and 2 are shown the magnetization, M , thermopower, S , and resistivity, ρ , as a function of temperature for the $x = 0$ and 0.5 samples. The difference in saturation moment between the samples is ascribed to the variation of compensation between the Fe and Cr sublattices as the Fe concentration is varied. The thermopower presented here agrees with previous measurements^{16,17}. For FeCr_2S_4 , $S(T)$ behaves as expected for an extrinsic p-type semiconductor with a decreasing Fermi energy below 300 K. The anomaly at $T_c = 170$ K bears some resemblance to that seen in the manganite perovskites, but, as it occurs in a semiconductor, most probably has a different origin and has previously ascribed to magnon-hole interactions¹⁷. For $\text{Fe}_{0.5}\text{Cu}_{0.5}\text{Cr}_2\text{S}_4$, $S(T)$ is n-type and also too large to be of metallic origin.

Also shown in Figs 1 and 2 is $\rho(T)$ for both samples at several different values of applied field. We note that the magnitude of $\rho(T)$ is about a factor of ten greater than found in vapour-transport-grown single crystals, but is comparable to that found in the same crystals after vacuum annealing¹⁸, implying possible sulphur

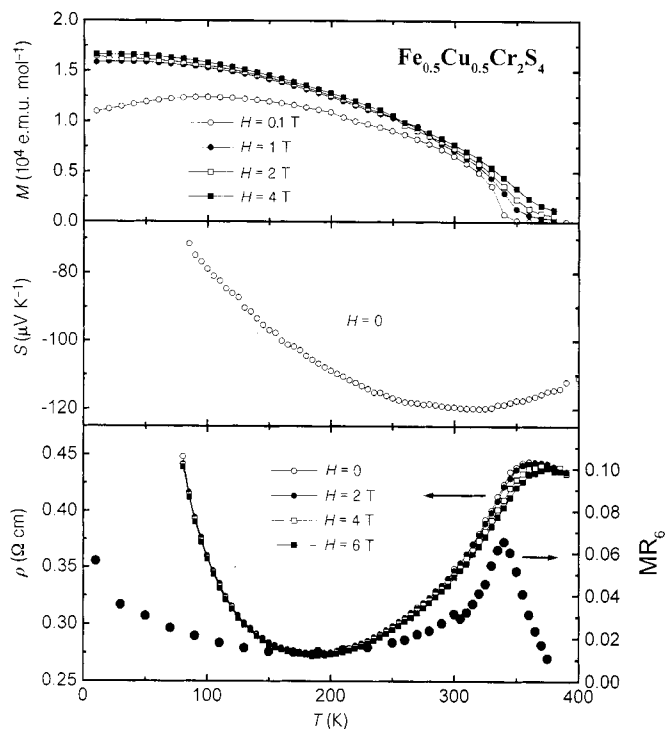


Figure 2 Magnetization (M), thermopower (S), resistivity (ρ) and magnetoresistance ($MR_H = (\rho(0) - \rho(H))/\rho(0)$), versus temperature for $\text{Fe}_{0.5}\text{Cu}_{0.5}\text{Cr}_2\text{S}_4$.

off-stoichiometry in our samples. We cannot rule out significant grain-boundary scattering in these polycrystalline samples, however. The low-field behaviour of $\rho(T)$ agrees with earlier measurements¹⁹. At temperatures above T_c , $\rho(T)$ obeys an activated form for FeCr_2S_4 ($\rho = \rho_0 e^{\Delta/T}$, $\Delta = 540$ K). Our limited measurement range does not allow a fit for $\text{Fe}_{0.5}\text{Cu}_{0.5}\text{Cr}_2\text{S}_4$ but clearly $d\rho/dT < 0$ for $T > T_c$, consistent with semiconducting behaviour in the paramagnetic state. At T_c , there is both an abrupt increase in $M(T)$ as T is lowered and a sharp drop in $\rho(T)$, similar to that seen in the manganite perovskites. Unlike in the perovskites, however, $\rho(T)$ never approaches typical metallic values for a half-filled band ($< 1 \text{ m}\Omega \text{ cm}$) despite the large region over which $d\rho/dT > 0$. Also shown in Figs 1 and 2 is the magnetoresistance, $MR_H(T) = (\rho(0) - \rho(H))/\rho(0)$, for $H = 6$ T, versus temperature. Similar to that found at low fields ($H < 1.2$ T; refs 18, 20) $MR_6(T)$ displays a peak at T_c , of a size comparable to that found in manganite perovskites.

To ascertain whether the $MR_6(T)$ peak is due to domain wall motion associated with the onset of FM order, or is intrinsic as in the perovskites, we measured the field dependence of ρ . In Figs 3 and 4 are shown high-field resistance data for FeCr_2S_4 and $\text{Fe}_{0.5}\text{Cu}_{0.5}\text{Cr}_2\text{S}_4$ respectively, at several different fixed temperatures spanning T_c (170 K and 340 K respectively). The dominant behaviour is large negative MR which does not saturate even at the highest fields measured. Thus it appears that the large MR is intrinsic but further studies on single crystals are needed to prove this point. It is interesting to compare MR_5 for $\text{Fe}_{0.5}\text{Cu}_{0.5}\text{Cr}_2\text{S}_4$ and the perovskite material $\text{La}_{0.67}\text{Sr}_{0.33}\text{MnO}_3$. From ref. 21, at $T/T_c = 0.8$, $MR_5 = 0.96$ for a single crystal of $\text{La}_{0.67}\text{Sr}_{0.33}\text{MnO}_3$. For the spinel, at the same reduced temperature, $MR_5 = 0.98$ which compares favourably with the perovskite oxide material. It would be of interest to compare MR of these materials at T_c . In Figs 3 and 4 we also show $M(H)$ below and near the MR peak temperature for both samples. Similar to the perovskites and according to expectations, $M(H)$ saturates at higher fields the closer T is to T_c . In the insets of Figs 3 and 4 are plotted the change in resistivity, $\Delta\rho = 1 - (\rho/\rho_{H=0})$, versus $(M/M_5)^2$ where M_5 is magnetization at 5 T

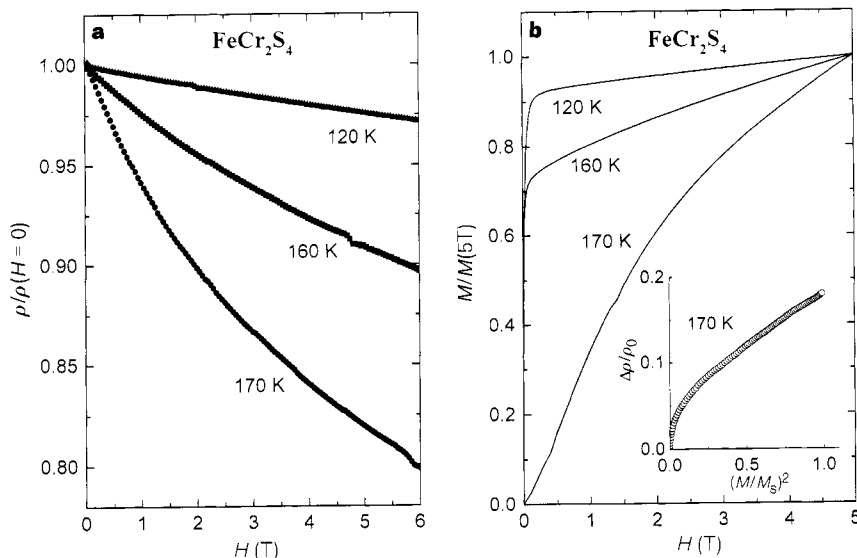


Figure 3 a, Resistivity, normalized to the value at $H = 0$, versus field for FeCr_2S_4 at several temperatures below T_c . **b**, Magnetization, normalized to the value at 5 T,

versus field several temperatures below T_c . Inset, $\Delta\rho/\rho_0$ versus $(M/M_s)^2$ for $T = T_c$. (M_s is the magnetization at 5 T).

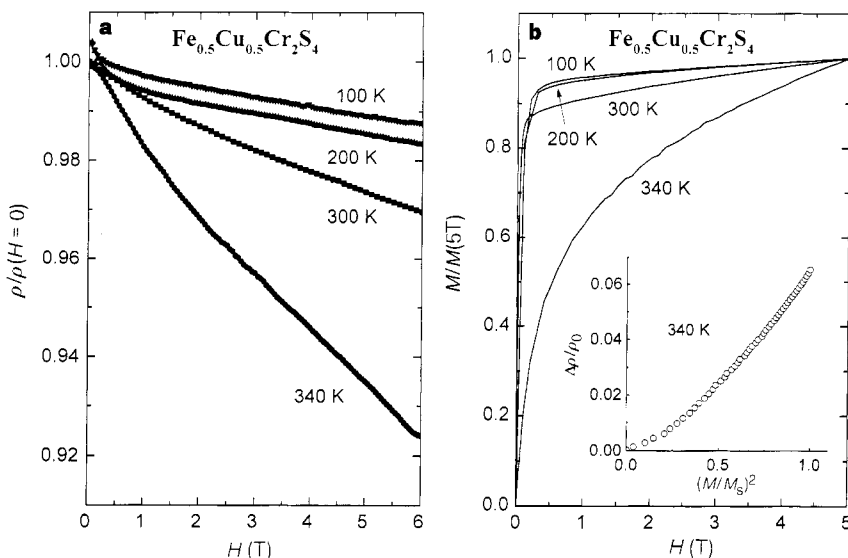


Figure 4 a, Resistivity, normalized to the value at $H = 0$, versus field for $\text{Fe}_{0.5}\text{Cu}_{0.5}\text{Cr}_2\text{S}_4$ at several temperatures below T_c . **b**, Magnetization, normalized

to the value at 5 T, versus field at several temperatures below T_c . Inset, $\Delta\rho/\rho_0$ versus $(M/M_s)^2$ for $T = T_c$.

for $T \approx T_c$. We note that above $H \approx 1$ T, there is a linear relationship, similar to that observed in $\text{La}_{0.825}\text{Sr}_{0.175}\text{MnO}_3$ (ref. 22).

The similarities in MR between the spinel compounds $\text{Fe}_{1-x}\text{Cu}_x\text{Cr}_2\text{S}_4$ and the perovskite compounds $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$ are somewhat surprising given the very different nature of electronic states responsible for conduction. Conduction in the CMR perovskites proceeds via hole-hopping among $\text{Mn}^{3+} - \text{Mn}^{4+}$ ($d^4 - d^3$) states at high temperatures and among a fully-spin-polarized itinerant band at low temperatures. In contrast, conduction in FeCr_2S_4 has been modelled as hopping among d^5 states above a valence band comprised of chalcogenide p -levels²³. The large change in $\rho(T)$ at T_c has been speculated to arise from scattering between holes and critical fluctuations¹⁷. A similarly large change in $\rho(T)$ occurs for CdCr_2Se_4 ($T_c \approx 110$ K) and here the idea of Anderson localization has been used to relate spin fluctuations in the magnetic subsystem to a mobility edge²⁴. It would be of interest to see if this idea is widely applicable in other low-carrier-density chalcogenides such as $\text{Fe}_{1-x}\text{Cu}_x\text{Cr}_2\text{S}_4$.

Another compound that exhibits CMR behaviour is $\text{Ti}_2\text{Mn}_2\text{O}_7$, which has the pyrochlore structure type²⁵⁻²⁷. This differs from the perovskites in that (1) there is no detectable amount of $d^3 - d^4$ mixing²⁷, (2) $\text{Ti}_2\text{Mn}_2\text{O}_7$ is metallic in the paramagnetic as well as in the ferromagnetic phases and (3) there is appreciable Ti-O character in the band structure near the Fermi level. Thus the Cr-chalcogenides represent a distinct third class of CMR systems which are semiconducting both at high ($T > T_c$) and low ($T \ll T_c$) temperature.

In $\text{Fe}_{1-x}\text{Cu}_x\text{Cr}_2\text{S}_4$, as in the manganese perovskites, the intrinsic MR at fields of order 100 Oe is less than 1%. This is to be compared to GMR multilayer materials where MR of order 10% at 100 Oe has been achieved in Permalloy/Au multilayers⁸. However, Hwang *et al.*²¹ have shown that the intrinsic low-field MR, as measured, for instance in single crystals of $\text{La}_{0.67}\text{Sr}_{0.33}\text{MnO}_3$, is substantially lower than found in polycrystalline samples of the same composition. This large low-field MR in polycrystalline samples has been attributed to spin-polarized tunnelling through grain boundaries.

Clocking transient chemical changes by ultrafast electron diffraction

J. Charles Williamson, Jianming Cao, Hyotcherl Ihee, Hans Frey & Ahmed H. Zewail

Arthur Amos Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, California 91125, USA

With the advent of femtosecond (fs) time resolution in spectroscopic experiments, it is now possible to study the evolution of nuclear motions in chemical and photobiochemical reactions. In general, the reaction is clocked by an initial fs laser pulse (which establishes a zero of time) and the dynamics are probed by a second fs pulse; the detection methods include conventional and photoelectron spectroscopy and mass spectrometry¹⁻⁴. Replacing the probe laser with electron pulses offers a means for imaging ultrafast structural changes with diffraction techniques⁵⁻⁸, which should permit the study of molecular systems of greater complexity (such as biomolecules). On such timescales, observation of chemical changes using electron scattering is non-trivial, because space-charge effects broaden the electron pulse width and because temporal overlap of the (clocking) photon pulse and the (probe) electron pulse must be established. Here we report the detection of transient chemical change during molecular dissociation using ultrafast electron diffraction. We are able to detect a change in the scattered electron beam with the zero of time established unambiguously and the timing of the changes clocked *in situ*. This ability to clock changes in scattering is essential to studies of the dynamics of molecular structures.

Diffraction methods, generally utilizing X-rays or electrons, are direct approaches for determining complex structures at microscopic dimensions⁹⁻¹⁴. Techniques for creating ultrashort, coherent and collimated X-ray pulses are still in their infancy, and at present the X-ray flux from most sources is relatively low. For gas-phase ultrafast electron diffraction (UED) to succeed, major challenges must be surmounted. First, because the number densities of gas samples are orders of magnitude lower than solids and surfaces, gas-phase diffraction intensities are very weak. Second, there is no long-range order to enhance coherent interferences, and thus the incoherent background scattering from gases is orders of magnitude greater. Last, there has not previously been a way to determine *in situ* the zero of time in UED experiments.

In gas-phase electron diffraction (GED) the scattering intensity decreases as the fourth power of the momentum transfer $s = (4\pi/\lambda)\sin(\theta/2)$, where λ is the de Broglie wavelength of the electrons and θ is the scattering angle. Conventional GED experiments require a high electron flux (of the order of microamperes) to yield internuclear separations with precision on the order of 0.001 Å (ref. 15). Over the past two decades, the speed of GED experiments has improved to the micro- and nanosecond timescale. Microsecond studies include time-of-flight investigations of molecular clusters¹⁶, and photodissociation experiments where the electron beam is electromagnetically chopped^{17,18}. A temporal resolution of 15 ns has been achieved with a pulsed laser driven electron source¹⁹.

In order to probe ps and sub-ps changes and the associated orientational anisotropy²⁰ with UED, three basic experiments must be incorporated into our apparatus: measurement of the electron pulsewidth, accurate clocking of the reaction, and detection of single electrons (necessary to reduce the electron flux and minimize space-charge broadening). To this end we have developed the new apparatus shown in Fig. 1, representative of a second generation of

This suggests the use of artificial nanostructures in CMR materials to achieve low-field MR in materials possessing intrinsic high-field MR. The present results on Cr-chalcogenides show that grain boundaries play less of a role than in the polycrystalline Mn-perovskite samples. This view is supported by the tendency of Cr-chalcogenides to form single crystals and the resultant high porosity of the present specimens. It would be of interest to increase the grain-boundary effect in the present system, either by varying the bulk synthesis route, or through thin-film techniques. □

Methods

The polycrystalline samples studied here were made by standard solid-state synthesis methods. Iron powder (99.9%), chromium powder (99.9%), sulphur powder (99.9995%) and copper powder (99.99%) were mixed in 2-g batches in the stoichiometric ratios and sealed into evacuated quartz tubes. The initial temperature was 450 °C, with slow increases in temperature in 50 °C increments up to 850 °C over a period of one week, with intermediate shakings and/or grinding until all signs of sulphur vapour and metal powder were eliminated. The resulting powder was then densified by pressing into pellets, (1/4 inch-diameter, 1/8 inch thick), re-sealed in a quartz tube and heated at 950 °C for three days. Phase purity was verified by X-ray powder diffraction of a small piece of the resultant pellet. Magnetization measurements were taken with a commercial magnetometer (Quantum Design Corp., San Diego, CA). Resistivity measurements were made using a four-probe in-line technique using an a.c. resistance bridge operating at 17 Hz. Thermopower data were taken using a commercial Seebeck-effect apparatus (MMR Corp., Mountain View, CA) referenced to a constantan standard.

Received 29 November 1996; accepted 7 February 1997.

- Jonker, G. H. & Van Santen, J. H. Ferromagnetic compounds of manganese with perovskite structure. *Physica* **16**, 337–349 (1950).
- Jin, S. *et al.* Thousandfold change in resistivity in magnetoresistive La-Ca-Mn-O films. *Science* **264**, 413–415 (1994).
- Von Helmolt, R. *et al.* Giant negative magnetoresistance in perovskitelike $\text{La}_{1-x}\text{Ba}_x\text{MnO}_3$ ferromagnetic films. *Phys. Rev. Lett.* **71**, 2331–2333 (1993).
- Chahara, K. *et al.* Magnetoresistance in magnetic manganese oxide with intrinsic antiferromagnetic spin structure. *Appl. Phys. Lett.* **63**, 1990–1992 (1993).
- Kusters, R. M. *et al.* Magnetoresistance measurements on the magnetic semiconductor $\text{Nd}_{0.5}\text{Pb}_{0.5}\text{MnO}_3$. *Physica B* **155**, 362–365 (1989).
- Millis, A. J., Shraiman, B. I. & Mueller, R. Dynamic Jahn-Teller effect and colossal magnetoresistance in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 175–178 (1996).
- Röder, H., Zhang, J. & Bishop, A. R. *Phys. Rev. Lett.* **76**, 1356–1359 (1996).
- Parkin, S. S. P. Giant magnetoresistance. *Annu. Rev. Mater. Sci.* **25**, 357–388 (1995).
- Hwang, H. Y., Cheong, S.-W. & Batlogg, B. Enhancing the low-field magnetoresistive response in perovskite manganites. *Appl. Phys. Lett.* **68**, 3494–3496 (1996).
- Sun, J. Z. *et al.* Observation of large low-field magnetoresistance in trilayer perpendicular transport devices made using doped manganite perovskites. *Appl. Phys. Lett.* **69**, 3266–3268 (1996).
- Lu, Y. *et al.* Large magnetotunnelling effect at low magnetic fields in micrometer-scale epitaxial $\text{La}_{0.67}\text{Sr}_{0.33}\text{MnO}_3$ tunnel junctions. *Phys. Rev. B* **54**, 8357–8360 (1996).
- Balters, P. K., Wojtowicz, P. J., Robbins, M. P. & Lopatin, E. Exchange interactions in ferromagnetic chromium chalcogenide spinels. *Phys. Rev.* **151**, 367–377 (1966).
- Goldstein, L. & Gibart, P. in *Proc. 17th Annu. Conf. on Magnetism and Magnetic Materials* (eds Graham, C.D. Jr. & Rhyne, J. J.) 883–886 (AIP Conf. Proc. No. 5 Am. Inst. Phys., New York, 1972).
- Amith, A. & Günsalus, G. L. Unique behavior of Seebeck coefficient in n-type CdCr_2Se_4 . *J. Appl. Phys.* **40**, 1020–1022 (1969).
- Haacke, G. & Beegle, L. C. Magnetic properties of the spinel system $\text{Fe}_{1-x}\text{Cu}_x\text{Cr}_2\text{S}_4$. *J. Phys. Chem. Solids* **28**, 1699–1704 (1967).
- Haacke, G. & Beegle, L. C. Chalcogenide spinels. *J. Appl. Phys.* **39**, 656–657 (1968).
- Haacke, G. & Beegle, L. C. Anomalous thermoelectric power of FeCr_2S_4 near the Curie temperature. *Phys. Rev. Lett.* **17**, 427–428 (1966).
- Watanabe, T. Electrical properties of FeCr_2S_4 and CoCr_2S_4 . *Solid State Commun.* **12**, 355–358 (1973).
- Watanabe, T. & Nakada, I. Preparation of some chalcogenide spinel single crystals and their electronic properties. *Jpn. J. Appl. Phys.* **17**, 1745–1754 (1978).
- Ando, K., Nishihara, Y., Okuda, T. & Tsushima, T. Hall effect and magnetoresistance in $\text{Fe}_{1-x}\text{Cu}_x\text{Cr}_2\text{S}_4$. *J. Appl. Phys.* **50**, 1917–1919 (1979).
- Hwang, H. Y., Cheong, S.-W., Ong, N. P. & Batlogg, B. Spin-polarized intergrain tunneling in $\text{La}_{0.5}\text{Sr}_{0.5}\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 2041–2044 (1996).
- Urushibara, A. *et al.* Insulator-metal transition and giant magnetoresistance in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. B* **51**, 14103–14109 (1995).
- Lotgering, F. K., Van Staple, R. P., Van Der Steen, G. H. A. M. & Wieringen, J. S. Magnetic properties of conductivity and ionic ordering in $\text{Fe}_{1-x}\text{Cu}_x\text{Cr}_2\text{S}_4$. *J. Phys. Chem. Solids* **30**, 799–804 (1969).
- Kogan, E. M. & Auslender, M. I. Anderson localization in ferromagnetic semiconductors due to spin disorder. *Phys. Stat. Sol. B* **147**, 613–620 (1988).
- Shimakawa, Y., Kubo, Y. & Manako, T. Giant magnetoresistance in $\text{Ti}_2\text{Mn}_2\text{O}_7$ with the pyrochlore structure. *Nature* **379**, 53–55 (1996).
- Cheong, S.-W., Hwang, H. Y., Batlogg, B. & Rupp, L. W. Jr. Giant magnetoresistance in pyrochlore $\text{Ti}_2\text{In}_2\text{Mn}_2\text{O}_7$. *Solid State Commun.* **98**, 163–166 (1996).
- Subramanian, M. A. *et al.* Colossal magnetoresistance without $\text{Mn}^{3+}/\text{Mn}^{4+}$ double exchange in the stoichiometric pyrochlore $\text{Ti}_2\text{Mn}_2\text{O}_7$. *Science* **273**, 81–84 (1996).

Correspondence and requests for materials should be addressed to A.P.R. (e-mail: apr@physics.bell-labs.com).

diffraction experiments in our laboratories^{6,8}. The apparatus is composed of a fs laser, an ultrafast electron gun, a free-jet expansion source, and a newly designed two-dimensional single-electron detection system. Femtosecond laser pulses were created from a colliding-pulse mode-locked ring dye laser. The output from this laser was directed through a four-stage pulsed dye amplifier with no pulse compression (620 nm, 2–3 mJ, 30 Hz, 300 fs). For initiation of the reaction, 95% of this beam was doubled (310 nm, ~250 μ J). The remainder (also doubled) was focused onto a back-illuminated, negatively-biased photocathode to generate the electron pulse via the photoelectric effect.

The ejected photoelectrons were accelerated into a series of electrostatic lenses that collimate and focus the electron beam; for 19 keV, λ is 0.088 Å. The position of the electron beam was controlled with horizontal and vertical deflectors. Two closely spaced metal plates were used for streaking experiments. The vacuum system was divided into two chambers and differentially pumped (Fig. 1). The sample molecules entered the vacuum chamber in a free-jet expansion mounted on an xyz positioning stage. The scattering chamber background pressure with gas flow was $\sim 2 \times 10^{-4}$ torr. Both chambers were shielded with μ -metal to reduce the effects of the Earth's magnetic field.

The component most critical to the success of UED is the detection system. The electron flux must be very low in order to keep the temporal resolution ultrafast. Early on we recognized that all of the scattered electrons must be detected for the experiment to succeed⁶, and we introduced the two-dimensional charge-coupled

device (CCD) as a detector in a direct electron bombardment mode⁸. To increase the longevity and flexibility of the single-electron detection system, our second-generation apparatus now employs two CCDs: one is a small, direct-bombardment device installed in the scattering chamber, and the other is a large, scientific-grade device mounted in a separate chamber at the end of a phosphor scintillator/fibre optic/image intensifier chain. Details will be published elsewhere.

To confirm the time resolution of the apparatus, we have measured the electron pulse duration *in situ* as a function of the number of electrons per pulse using the high-speed deflection technique in a streak camera arrangement. Streak velocities (which measure the conversion from temporal to spatial resolution, as shown in Fig. 2) attained with this setup are on the order of 1.6×10^8 m s⁻¹, or ~2 pixels per ps on the detector. The resolution of the streak experiment is therefore ± 0.5 ps. Steak images spanning three orders of magnitude of electron number are shown in Fig. 2.

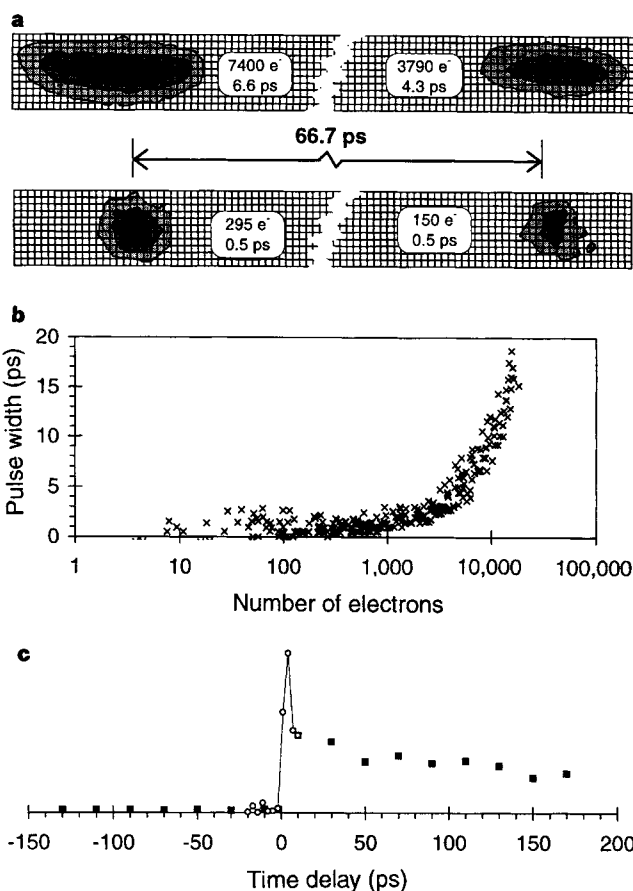


Figure 2 a, Surface contour plots of low- and high-intensity streaked electron pulse pairs. Each pair comes from two incident light pulses, one of which is delayed by 66.7 ps. The horizontal coordinate is the streaking axis. The temporal durations of pulses with high electron numbers (top trace: 7,400 e⁻, 6.6 ps; 3,790 e⁻, 4.3 ps) are broadened by space-charge effects. For low electron numbers (bottom trace: 295 e⁻, 0.5 ps; 150 e⁻, 0.5 ps), measurement of the pulse width is limited by the resolution of the streak experiment. The electron pulse width was obtained by subtracting the width of the unstreaked electron beam (4 pixels) from the width of the streaked pulses (lorentzian profiles). **b**, Measured electron pulse widths as a function of the number of electrons. The streak velocity was 2.1 pixels per ps, or 1.6×10^8 m s⁻¹. The sweep voltage is ~1 kV ns⁻¹. **c**, Experimental photoionization-induced lensing transient. The ordinate reflects the relative difference between the average profile of the electron beam spot in the presence of the excitation laser and a reference profile (no laser).

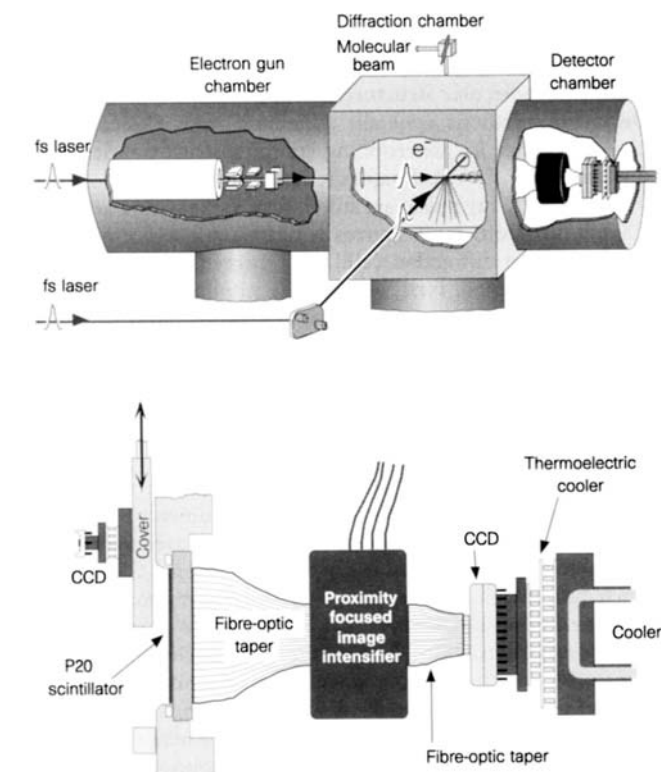


Figure 1 Top, second-generation ultrafast electron diffraction apparatus, consisting of an electron gun chamber, a diffraction chamber and a detector chamber. Two fs laser pulses are used; the first initiates the chemical change and the second generates the electron pulse. Bottom, detection scheme. Incident electrons either directly bombard a small CCD or strike a phosphor-coated fused fibre-optic window. Light emitted from the phosphor is amplified by an image intensifier and brought to a scientific-grade CCD. Both CCDs are thermoelectrically cooled.

For electron intensities on the order of 10^4 electrons per pulse, the width is typically ~ 10 ps. The laser and electron beam waists at overlap was measured to be $450 \pm 50 \mu\text{m}$, so the velocity mismatch contribution⁷ to the total experimental temporal resolution is ~ 4 ps.

To clock chemical changes, we must establish the zero of time (t_0). A streak camera⁹ or photo-triggered deflection plates²¹ can give t_0 to within 10–20 ps, but these methods require some extrapolation and do not duplicate the design of crossed-beam experiments. Another approach is to rely on changes in the diffraction pattern of the

system under investigation, but this is not an independent means of finding t_0 . More importantly, this method is simply not practical for gas-phase work because of the long integration times required to obtain a single data point.

In the clocking technique developed here for UED, we use the crossed-beam geometry of the actual diffraction experiment, rather than relying on scattered electrons. The idea is to let the initiation laser pulse create a coulombic field (by ionization), which deflects the unscattered electron beam only during and after the laser pulse passes through. The phenomenon, termed lensing, occurs because the field focuses the charged electron beam. Fig. 2 shows the degree of lensing versus time; the lensing is a maximum when the laser and electron pulses are temporally overlapped. The results accurately show the precise $t = 0$ (to ~ 2 ps) and hence allow a direct clocking of changes in the diffraction experiment with ps resolution. We recorded lensing transients as a function of the position, polarization and diameter of the pump laser, and conducted control experiments where no gas was flowing and where the sample source tip was far removed from or close to the interaction region.

After successfully establishing t_0 , we conducted the first time-resolved gas-phase electron diffraction experiment with picosecond resolution. Diiodomethane (CH_2I_2) was selected as our prototype system because the loss of an iodine atom after dissociation at 310 nm is a major structural change which occurs in less time than the rotational period (~ 10 ps)^{22,23}. To clock the change, diffraction images were recorded on the CCD at delays of -20 ps, 0 ps, and up to $+70$ ps (Figs 3 and 4).

The experimental modified molecular scattering intensity, $sM(s)$, was obtained at each time point

$$sM(s) = s \frac{I_{\text{tot}}(s) - I_{\text{back}}(s)}{|f_a| |f_b|} \quad (1)$$

where I_{tot} is the total scattering intensity, I_{back} is the background (including atomic) scattering intensity, and f_a and f_b are atomic scattering amplitudes^{15,19}. Typical $sM(s)$ raw data taken at -20 ps

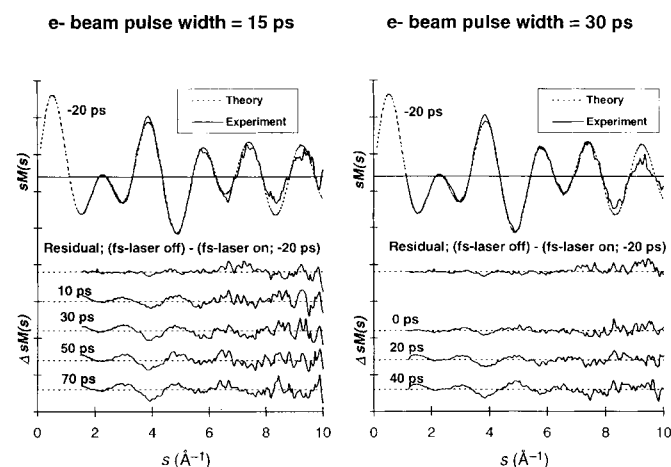


Figure 3 Experimental modified molecular scattering curves for two sets of data taken with 15-ps (left panel) and 30-ps (right panel) electron pulses. The delay times between the fs laser pulse and the ps electron pulse are shown. Theoretical calculations of $sM(s)$ at -20 ps are presented (see text).

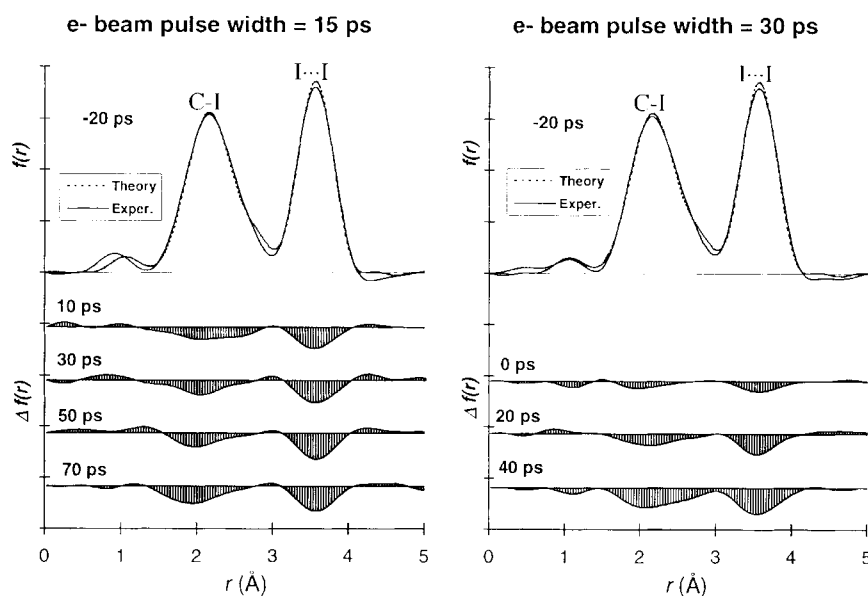


Figure 4 Radial distribution functions $f(r)$ and $\Delta f(r)$ obtained from the $sM(s)$ and $\Delta sM(s)$ curves of Fig. 3, also for two sets of data (15-ps and 30-ps electron pulse width, left and right panel, respectively). The corresponding theoretical $f(r)$ curve for the -20 ps data is shown. The changes obtained are in the region of the C-I and I...I internuclear separations (see text). Note that in $f(r)$ the peak corresponding to the I...I distance should have about twice the area of the first peak (C-I and H...I); the peak area scales approximately as $n_i Z_i Z_j / r_{ij}$ (where Z is the atomic

number and n is the multiplicity for nuclear pairs). However, because of damping in equation (2), the peaks appear to be of equal areas, and this is evident in our simulations for different values of k . The theoretical simulations shown in the figure are based on the structural data of ref. 24. We have also used more recent data from K. Hedberg's and co-workers (personal communication) and obtained very similar $f(r)$ curves (data not shown), within the reported resolution.

are shown in Fig. 3. The theoretical fits, calculated from structural parameters²⁴ (see also Fig. 4 legend) of CH₂I₂, are superimposed on the experimental data. To unambiguously establish the magnitude of the change with time, we also recorded the *sM(s)* in the absence of the fs initiation laser and obtained the residual of that *sM(s)* minus the *sM(s)* taken at -20 ps. This residual is essentially flat (Fig. 3), indicating that the two are identical. Thus, the *sM(s)* at -20 ps is a calibration point in time, independent of any theoretical fit. The change in the diffraction pattern as a function of time is shown in two sets of data having differing temporal resolutions (15 ps and 30 ps); the $\Delta sM(s)$ curves are relative to the -20 ps data.

Radial distribution curves were obtained from the *sM(s)* curves according to the standard GED equation:

$$f(r) = \int_0^{s_{\max}} sM(s) \sin(sr) \exp(-ks^2) ds \quad (2)$$

where the constant *k* is a damping coefficient included for the limited *s* range. Figure 4 shows the experimental and theoretical *f(r)* obtained from the -20 ps *sM(s)* curves of Fig. 3. The time-evolution of $\Delta f(r)$ is also presented for the two sets of data. At positive time the amplitudes of the dominating C-I and I...I peaks decrease, corresponding to the loss of an iodine atom. Both the *sM(s)* and *f(r)* representations of our diffraction data clearly show a chemical change occurring as the delay between the fs laser pulse and the ps electron pulse sweeps from negative, to zero, and to positive times. From the *f(r)* data, we obtained the standard deviation for more than 100 diffraction images for each data set, and established that the signal-to-noise ratio is better than 60. Thus, we could detect a change as small as 3%, which is a factor of 5 less than the changes reported here.

Breakage of one C-I bond reduces the I...I peak twice as quickly as the C-I peak because CH₂I₂ has two C-I bonds and one I...I internuclear separation. This is observed in our data, corroborated by theoretical simulations of *f(r)* (not shown) which account for the percentage change of CH₂I₂ to CH₂I and I. Similar time-resolved studies were made for C₂F₄I₂ and CHI₃. Quantitative analysis in the future should yield the changes due to the second C-I bond breakage, and elucidate the structure of the intermediate. Last, ground-state diffraction patterns of three other molecules (CF₃I, CCl₄ and SF₆) were recorded with our new UED apparatus (data not shown). As mentioned earlier, the diffraction patterns were converted to experimental *sM(s)* curves and radial distribution functions. The experimental and theoretical agreement is very good.

From the results presented here, it is now evident that ultrafast electron diffraction has achieved the temporal resolution and detection sensitivity necessary for picosecond and subpicosecond studies of complex molecular systems. In future work, utilizing modulation and difference detection techniques, efforts will be directed toward the suppression of background scattering from unreactive species so that molecular structural changes can be analysed with precision. □

Received 8 October 1996; accepted 21 January 1997.

- Chergui, M. (ed.) *Femtochemistry: Ultrafast Chemical and Physical Processes in Molecular Systems* (World Scientific, Singapore, 1996).
- Manz, J. & Wöste, L. (eds) *Femtosecond Chemistry* (VCH, New York, 1995).
- Zewail, A. H. *Femtochemistry: Ultrafast Dynamics of the Chemical Bond* (World Scientific, Singapore, 1994).
- Manz, J. & Castleman, A. W. *Femtochemistry*, *J. Phys. Chem. (spec. iss.)* **97**(48), 12423–12649 (1993).
- Williamson, J. C. & Zewail, A. H. Structural femtochemistry: experimental methodology. *Proc. Natl Acad. Sci. USA* **88**, 5021–5025 (1991).
- Williamson, J. C., Dantus, M., Kim, S. B. & Zewail, A. H. Ultrafast diffraction and molecular structure. *Chem. Phys. Lett.* **196**, 529–534 (1992).
- Williamson, J. C. & Zewail, A. H. Ultrafast electron diffraction. Velocity mismatch and temporal resolution in crossed-beam experiments. *Chem. Phys. Lett.* **209**, 10–16 (1993).
- Dantus, M., Kim, S. B., Williamson, J. C. & Zewail, A. H. Ultrafast electron diffraction. 5. Experimental time resolution and applications. *J. Phys. Chem.* **98**, 2782–2796 (1994).
- Räksä, E. et al. Ultrafast x-ray absorption probing of a chemical reaction. *J. Chem. Phys.* **104**, 6066–6069 (1996).
- Tomov, I. V., Chen, P. & Rentzepis, P. M. Picosecond time-resolved x-ray diffraction during laser-pulse heating of an Au(111) crystal. *J. Appl. Crystallogr.* **28**, 358–362 (1995).

- Kim, K.-J., Chattopadhyay, S. & Shank, C. V. Generation of femtosecond X-rays by 90-degrees Thomson scattering. *Nucl. Instr. Meth. Phys. Res. A* **341**, 351–354 (1994).
- Williamson, S., Mourou, G. & Li, J. C. M. Time-resolved laser-induced phase transformation in aluminum. *Phys. Rev. Lett.* **52**, 2364–2367 (1984).
- Elsayed-Ali, H. E. Time-resolved reflection high-energy electron diffraction of metal surfaces. *Proc. SPIE* **2521**, 92–102 (1995).
- Aeschlimann, M. et al. A picosecond electron gun for surface analysis. *Rev. Sci. Instrum.* **66**, 1000–1009 (1995).
- Hargittai, I. & Hargittai, M. (eds) *Stereochemical Applications Of Gas-Phase Electron Diffraction* (VCH, New York, 1988).
- Dibble, T. S. & Bartell, L. S. Electron diffraction studies of the kinetics of phase changes in molecular clusters. 3. Solid-state phase transformations in SeF₆ and (CH₃)₃CCl. *J. Phys. Chem.* **96**, 8603–8610 (1992).
- Ischenko, A. A. et al. A stroboscopic gas-electron diffraction method for the investigation of short-lived molecular species. *Appl. Phys. B* **32**, 161–163 (1983).
- Rood, A. P. & Milledge, J. Combined flash-photolysis and gas-phase electron diffraction studies of small molecules. *J. Chem. Soc. Faraday Trans. 2* **80**, 1145–1153 (1984).
- Ischenko, A. A., Schäfer, L., Luo, J. Y. & Ewbank, J. D. Structural and vibrational kinetics by stroboscopic gas electron diffraction: the 193 nm photodissociation of CS₂. *J. Phys. Chem.* **98**, 8673–8678 (1994).
- Williamson, J. C. & Zewail, A. H. Ultrafast electron diffraction. 4. Molecular structures and coherent dynamics. *J. Phys. Chem.* **98**, 2766–2781 (1994).
- Williamson, S., Mourou, G. & Letzring, S. Picosecond electron diffraction. *Proc. SPIE* **348**, 313–317 (1983).
- Kawasaki, M., Lee, S. J. & Bersohn, R. Photodissociation of molecular beams of methylene iodide and iodoform. *J. Chem. Phys.* **63**, 809–814 (1975).
- Baughcum, S. L. & Leone, S. R. Photofragmentation infrared emission studies of vibrationally excited free radicals CH₂I and CH₂I₂. *J. Chem. Phys.* **72**, 6531–6545 (1980).
- Hassel, O. & Viervoll, H. Electron diffraction investigation of molecular structure II, results obtained by the rotating sector method. *Acta Chem. Scand.* **1**, 149–168 (1947).

Acknowledgements. We thank S. Preuss for his help in the construction and testing of the experimental apparatus, and H. Elsayed-Ali for providing us with part of the design of the electron gun.

Correspondence should be addressed to A.H.Z.

Three-dimensional self-assembly of millimetre-scale components

Andreas Terfort, Ned Bowden & George M. Whitesides

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

The spontaneous association of molecules, termed molecular self-assembly, is a successful strategy for the generation of large, structured molecular aggregates¹. The most important source of inspiration for this strategy is the biological world, in which many processes involve interfacial interactions and shape selectivity that guide the formation of complex, multicomponent three-dimensional structures. The success of molecular self-assembly notwithstanding, many objectives in science and technology require the assembly of components that are much larger than molecules: examples include microelectronic and microelectromechanical systems, sensors and microanalytical and micro-synthetic devices². Photolithography, the principal technique used to make such microstructures, has certain limitations: it cannot easily form non-planar or three-dimensional structures; it generates structures that are metastable; and it can be used only for a limited set of materials³. Here we describe an approach for the self-assembly of millimetre-scale components that uses two concepts to direct the assembly process: shape recognition and the minimization of liquid–liquid interfacial free energies⁴. These play a role in other spontaneous self-assembly phenomena, such as the formation of bubble rafts^{5,6}, the patterned dewetting of surfaces^{7,8}, and the coalescence of liquid drops⁹. We apply self-assembled monolayer molecular films¹⁰ to the surfaces of shaped macroscopic objects to render them hydrophilic or hydrophobic, depending on the terminal groups of the bound molecules. In aqueous solution, hydrophobic surfaces bearing a thin film of a

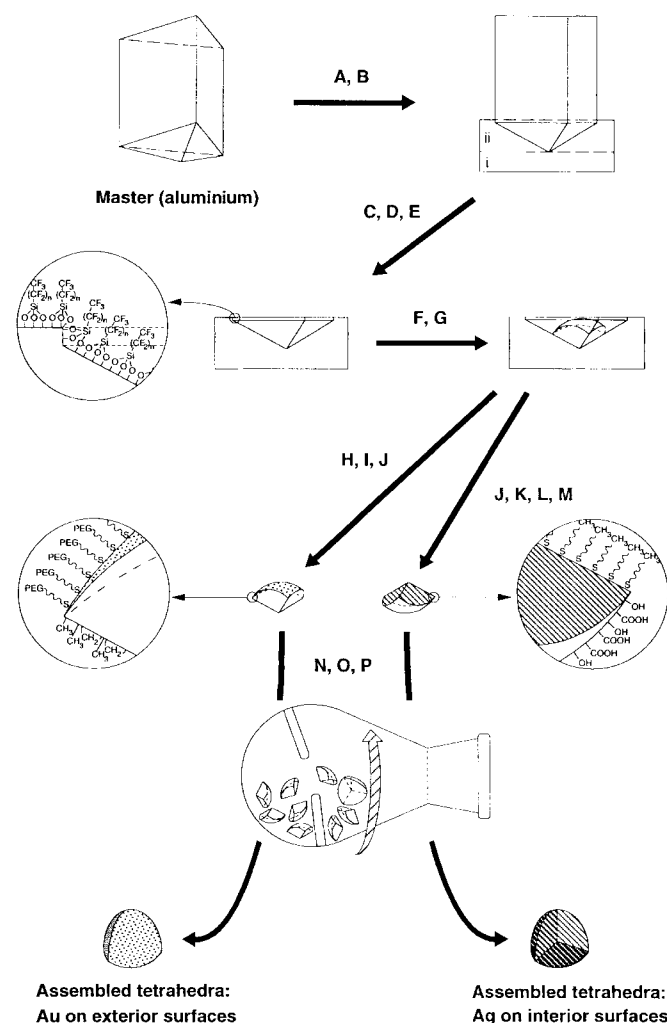


Figure 1 Schematic outline of the process used to fabricate and assemble components of the tetrahedra. The master (top left) was fabricated from aluminium by machining; the angles defining the edges at the tip were 109° . Process steps are as follows: (A) A pad of polydimethylsiloxane (PDMS, i) was cast and cured. (B) The master was cast into a second layer (ii) of PDMS. (C) After curing, the master was removed. (D) This PDMS mould was oxidized in an oxygen plasma to generate surface Si-OH groups. (E) Exposure to perfluoroalkyl trichlorosilane vapour resulted in the formation of a perfluoroalkylated surface (see inset). (F) A measured amount of the liquid polyurethane prepolymer was added by means of a syringe pump and (G) was cured in ultraviolet light. (H) Gold was evaporated on top of the parts while they remained in the mould. (I) A hydrophilic self-assembling monolayer was formed on the gold surface by immersion in a solution of a thiol terminating in polyethyleneglycol (PEG) groups (the PEG used was methoxy-terminated material with molecular weight 350); this thiolate is hydrophilic and resists adhesion. (J) The parts were removed from the mould. (K) Silver was evaporated on the bottom of the parts. (L) The parts were oxidized in an oxygen plasma. (M) A hydrophobic self-assembling monolayer was generated on the Ag surface by immersion of the parts in a solution of $\text{CH}_3(\text{CH}_2)_{16}\text{SH}$. (N) A number of parts (typically 20–100) were placed in the assembly flask and the liquid adhesive and ethanol were added and distributed. (O) Water was added until the flask was completely filled, and the mixture was rotated mechanically for several hours; the tetrahedra assembled themselves. (P) If the glue was photocurable, the parts ‘froze’ into place on ultraviolet irradiation. They were then removed and examined.

hydrophobic, lubricating liquid adhere to similar surfaces with complementary shapes, while being able to adjust their relative alignment to ensure a good fit. In this way, the components assemble into well defined aggregates, which can be bound permanently when the hydrophobic liquid films consist of a polymerizable adhesive.

Figure 1 outlines our process. The individual components to be assembled were formed from a hydrophobic polymer (typically a polyurethane: NOA-63 or NOA-88; Norland, New Brunswick, NJ) by moulding. The surface of these components was patterned into hydrophilic and hydrophobic regions using two procedures. In one, some portions of the surface were covered (often by leaving the units in the moulds in which they were made), and the exposed surfaces were oxidized with an oxygen plasma. This process left the hidden surfaces hydrophobic, while making the exposed surfaces hydrophilic. In a second, differentiation was achieved by coating the regions with gold or silver by evaporation, and assembling monolayers of hydrophobic or hydrophilic¹⁰ thiols onto them. The components were placed in a flask, their surface was wetted with ethanol, a small amount ($0.05\text{--}0.1\text{ }\mu\text{l per mm}^2$ of hydrophobic surface) of a hydrophobic liquid—either an alkane or a photopolymerizable adhesive, such as dodecyl methacrylate—was added, and the system was stirred gently to coat the surfaces. Water was added; the water displaced the hydrophobic liquid from the hydrophilic surfaces, while leaving a thin film of liquid coating the hydrophobic surfaces. The aqueous suspension of the components was tumbled in a rotating flask. This agitation brought the compo-

nents into contact. The hydrophilic surfaces stuck neither to other hydrophilic surfaces nor to the hydrophobic surfaces with their thin coating of liquid, whereas the hydrophobic surfaces adhered to one another. Once the surfaces were in contact, the liquid acted as a lubricant, allowing the components to adjust their position relative to one another by lateral movement. This movement minimized the interfacial free energy of the system by minimizing the surface area of the lubricant–water interface. After the systems had approached equilibrium, ultraviolet irradiation (when the liquid was a photopolymerizable adhesive) ‘froze’ the assembly into permanent form. These components were removed and examined.

Figure 2 shows small assemblies fabricated using this procedure. The most complicated of the structures formed was approximately a sphere: four pieces with hydrophobic faces defined by planes meeting at the tetrahedral angle (109°) were the components that assembled. A key component to these assemblies was to make the contacting surfaces smooth; if they were rough, the lateral motion required to reach the configuration with the maximum contacting surface was hindered by direct, unlubricated contact and adhesion of the surfaces. To generate suitably smooth surfaces, the faces of the aluminium master were polished to a mirror finish; we did not measure the surface roughness.

The assemblies in Fig. 2a–c are symmetrical. It was also possible to generate unsymmetrical assemblies by taking advantage of docking of unsymmetrical components using shape-selective recognition; Fig. 2d is a demonstration of the principle. The assembly of complementary shapes is integral to the formation of most

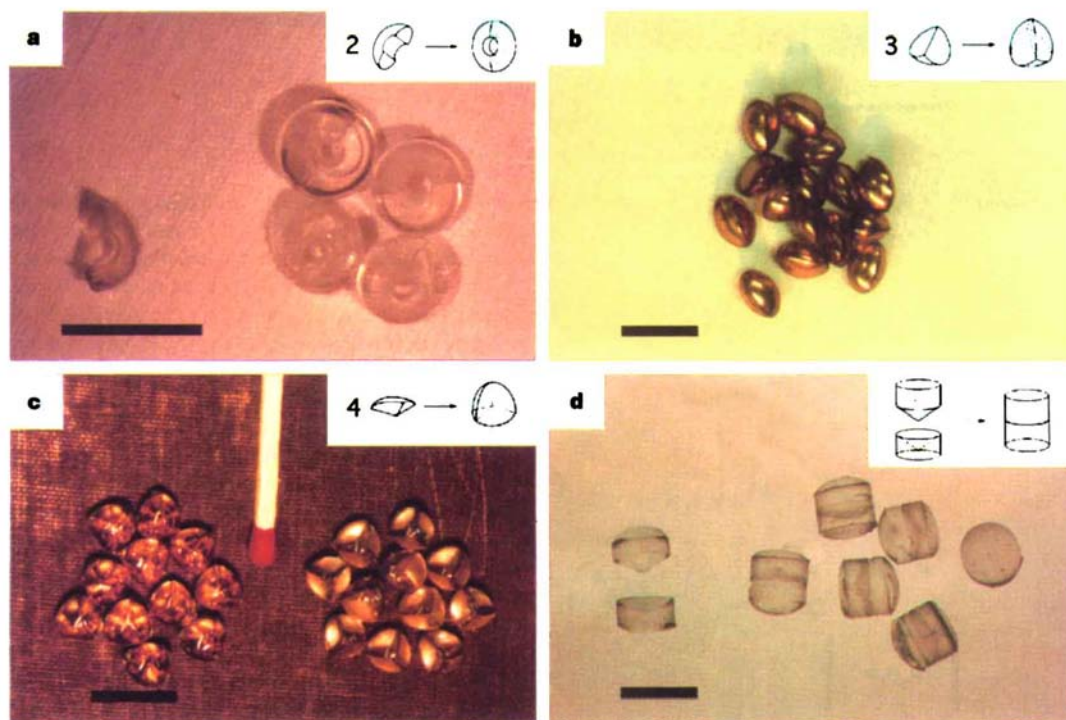


Figure 2 Several examples of self-assembled structures. Scale bars, 1 cm. **a**, Two arches (one of them at the left) form small 'doughnuts'. The differentiation was achieved in this case by oxidation of the outer sides, while the adhering surfaces were unmodified, hydrophobic polymer (NOA-63). **b**, Small 'footballs' formed from three identical parts. The adhesive sides were unmodified polymer (NOA-63) and the curved outer surface was a 20-nm-thick layer of gold (with an underlying 2-nm layer of titanium as adhesion promoter) covered with a self-assembling monolayer of a hydrophilic, adhesion-resistant polyethylene glycol terminated thiol. **c**, Tetrahedron-like structures formed by self-assembly of four identical

parts. The surface treatment for the objects on the left was as in **b** (gold on the exterior faces, see also Fig. 1). On the right, the adhesive, interior sides of the tetrahedra were covered with a hydrophobic self-assembling monolayer ($\text{CH}_3(\text{CH}_2)_{15}\text{SH}$) on silver, while the outer sides were plasma-oxidized polymer. This configuration made it possible to visualize the interior surfaces. **d**, Two different parts assembled in a lock-and-key manner. The adhering surfaces were unmodified polymer, while the outer sides were made hydrophilic by treatment with an oxygen plasma (see Fig. 1 for details of surface processing).

complex biological systems. Our demonstration indicates that the same strategy can be applied to non-biological assemblies.

This work establishes that shape-selective recognition of surfaces and minimization of interfacial free energies can be combined to provide a strategy for assembling small components into three-dimensional assemblies. Because association is reversible, the system is self-repairing; the experiments described here usually result in quantitative formation of correctly assembled objects. We believe that we will be able to assemble smaller structures using the same or a similar methodology, and that we will be able to build functional devices. □

Received 1 November 1996; accepted 3 January 1997.

1. Philp, D. & Stoddart, J. F. *Angew. Chem. Int. Edn Engl.* **35**, 1155–1196 (1996).
2. Harrison, D. J., Fluri, K., Chiem, N., Tang, T. & Fan, Z. in *Sensors and Actuators B Chemical Vol. B33* (eds Middelhoeck, S. & Cammann, K.) 105–109 (Elsevier, Amsterdam, 1996).
3. Moreau, W. M. *Semiconductor Lithography: Principles, Practices, and Materials* (Plenum, New York, 1988).
4. Kim, E. & Whitesides, G. M. *Chem. Mater.* **7**, 1257–1264 (1995).
5. Simpson, A. W. & Hodgkinson, P. H. *Nature* **237**, 320–322 (1972).
6. Schöber, S. T., Friederich, J. & Altmann, A. *J. Appl. Phys.* **71**, 2206–2210 (1992).
7. Gorman, C. B., Biebuyck, H. A. & Whitesides, G. M. *Chem. Mater.* **7**, 252–254 (1995).
8. Biebuyck, H. A. & Whitesides, G. M. *Langmuir* **10**, 2790–2793 (1994).
9. Blass, E. *Chem.-Ing.-Tech.* **60**, 935–947 (1988).
10. Pale-Grosdemange, C., Simon, E. S., Prime, K. L. & Whitesides, G. M. *J. Am. Chem. Soc.* **113**, 12–20 (1991).

Acknowledgements. This work was supported by the US ONR/DARPA and the US NSF. A.T. thanks the Deutsche Forschungsgemeinschaft for a research grant; N.B. was supported by a predoctoral fellowship from the US Department of Defense.

Correspondence and requests for materials should be addressed to G.M.W. (e-mail: gw@whiteides@gmgroup.harvard.edu).

Spurious trends in satellite MSU temperatures from merging different satellite records

James W. Hurrell & Kevin E. Trenberth

PO Box 3000, National Center for Atmospheric Research, Boulder, Colorado 80307, USA

Analysis of global surface air temperature records has indicated that recent years have been among the warmest since the late nineteenth century¹, with 1995 being the warmest year on record². But the rate of global annual mean surface warming of 0.13 °C per decade during the period 1979–95 differs substantially from the global lower-tropospheric cooling trend of –0.05 °C per decade³ inferred from the record (MSU-2R) of radiance measurements by the satellite Microwave Sounder Unit (MSU)^{4,5}. Accordingly, the satellite record has been widely cited by sceptics as evidence against global warming^{6–10}. However, a substantial fraction of the measured radiance originates not from the atmosphere but from the Earth's surface¹¹, and gives rise to high noise levels. This noise can lead to errors when merging temperature time series

obtained from different satellites. Here we present comparisons among different MSU retrievals, sea surface temperatures (SSTs), and equivalent MSU temperatures derived from an atmospheric general circulation model forced with observed SSTs. The comparisons, focused on the tropics where atmospheric temperatures are closely tied to SSTs, strongly suggest that two spurious downward jumps occur in the MSU-2R record coinciding with changes in satellites, and that the real trend in MSU temperatures is likely to be positive, albeit small.

The relative merits of the surface and MSU temperature records have been a matter of spirited debate¹. One often overlooked issue is that there is no single satellite record, and that different tropospheric measures of temperature from the MSUs contain different trends and different error characteristics. The trend since 1979 in global departures from the mean annual cycle (henceforth 'anomalies') from MSU channel 2 (MSU-2), which is representative of the middle troposphere, is $+0.02^{\circ}\text{C}$ per decade. Although the MSU-2 data are more reliable than the MSU-2R record, for reasons discussed below, they contain a non-trivial influence from the stratosphere which can be removed through a linear combination of the off-nadir data. The result is a narrower vertical weighting function (MSU-2R) which peaks lower in the troposphere but is more sensitive to surface effects⁵. The MSU-2R data have been used in most of the more recent comparisons to surface air temperatures¹.

Another important factor is that the surface and MSU records measure different physical quantities so that decadal trends should not be expected to be the same^{12,13}, especially in the presence of strong interannual variability associated with volcanic eruptions and El Niño/Southern Oscillation (ENSO) events which affect the vertical temperature structure^{14,15}. Unreconciled discrepancies among the two records remain^{12,15}, however, and are especially

evident in the tropics (20°S – 20°N) in regions where there are few reliable radiosonde temperature records.

From 1979 to 1995, the linear trend in MSU-2R temperature anomalies averaged over the tropics is -0.11°C per decade, in contrast to the surface warming rate of 0.10°C per decade. Physical differences between the two measures of temperature cannot fully account for this discrepancy¹². The linear trend in tropical MSU-2 anomalies since 1979 is 0.05°C per decade, which is closer to the observed rate of warming in the surface record and is identical to the linear trend in the SSTs. There is a strong cooling trend in the MSU-2R anomalies after the effects of the SST variations are removed using linear regression (Fig. 1). The regression is based on the period 1982–91 when there is good agreement between the two records in spite of the large ENSO-related variations, and the variance explained is greatest with SST leading by one month if a single lag is used. Reconstructed SST analyses¹⁶ are used through 1981 with optimally interpolated SST analyses¹⁷ thereafter, but there is virtually no sensitivity of results to alternative SST analyses (for example, based solely on *in situ* data from the Comprehensive Ocean Atmosphere Dataset). MSU-2R anomalies are 0.25°C warmer than expected from the SSTs before 1982 and are colder after 1991 by 0.1°C . The MSU-2 residual is also positive before 1982, but not as much as the MSU-2R residual, and it does not show the nearly stepwise cooling evident in the MSU-2R data in mid-1991. Moreover, the MSU-2 record warms relative to MSU-2R over the tropics with differences (MSU-2R – MSU-2) of 0.16°C averaged over 1979–80 and -0.13°C over 1992–95 (Fig. 1). The differences in trends arise largely from differences over tropical land where the 17-yr MSU-2R (MSU-2) trend is -0.24 (0.01) $^{\circ}\text{C}$ per decade compared with -0.07 (0.07) $^{\circ}\text{C}$ per decade over the oceans.

Here we use an atmospheric general circulation model forced

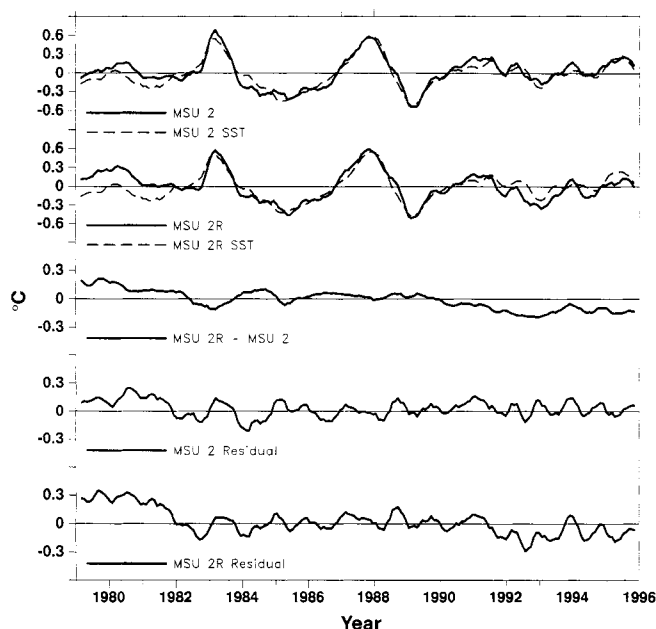


Figure 1 Five-month running mean MSU-2 and MSU-2R temperature anomalies and the differences (MSU-2R – MSU-2) area-averaged over the tropics (20°S – 20°N). The dashed lines represent the MSU anomalies associated with sea surface temperature (SST) variations derived from linear regression with a one-month lead in SST over the period 1982–91. Also shown are the residuals after removing the linear SST effects for MSU-2 and MSU-2R.

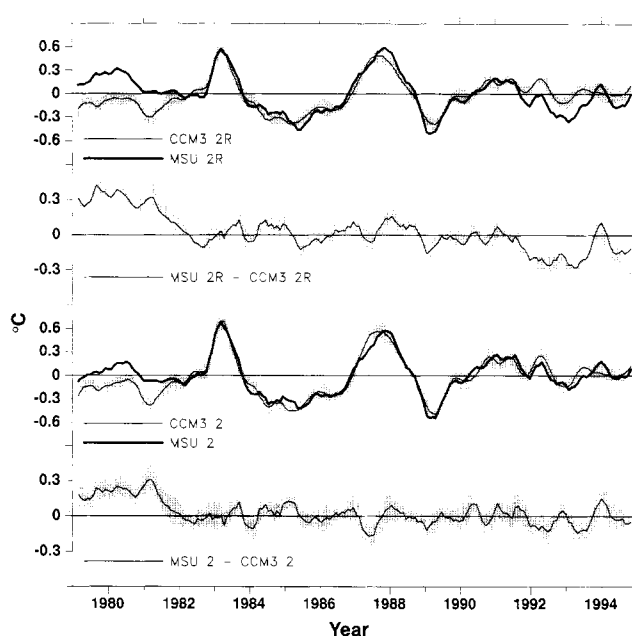


Figure 2 Five-month running mean MSU-2 and MSU-2R temperature anomalies (thick lines) and the ensemble mean (light lines) plus range about the mean (shading) simulated MSU anomalies from the CCM3 integrations area-averaged over the tropics (20°S – 20°N). Also shown are the differences (MSU – CCM3) for MSU-2R and MSU-2 (bottom).

with observed SSTs to nonlinearly translate the SST record into a temperature record equivalent to that of the MSUs. Version 3.0 (CCM3) is the fourth generation in the series of NCAR Community Climate Models¹⁸ and is run at wavenumber 42 (T42) resolution (the transform grid is 2.8°) with 18 vertical levels. The last 16 years of a five-member ensemble of 45-yr integrations forced with observed monthly SSTs are analysed. Over the decade 1982–91, MSU-2R anomalies lie within the CCM3 spread with only a few minor exceptions even though there are large ENSO-related excursions, and the correlation with the ensemble mean is 0.92 (Fig. 2). Over the entire record, however, the correlation falls to 0.73 because of large discrepancies during 1979/80 and after 1991. For MSU-2, simulated anomalies compare more closely with the satellite record, although the MSU-2 temperatures are warmer early in the record. Moreover, simulated MSU-2R and MSU-2 anomalies vary together, so there is no significant difference in trends, unlike the MSU data.

It should be recognized that anomalous forcings of the atmosphere other than from SSTs took place during this period. In particular, the cold anomalies at the end of the record may be partly attributable to the direct (atmospheric) effect of the Mt Pinatubo eruption (June 1991)^{14,15}, but differences of similar magnitude have persisted over the past several years when the effects of the eruption should have diminished. Thus it appears that most of the effects from Pinatubo are captured by the SST changes.

The above discrepancies need to be explained; one important issue is how the records from different satellites are merged. The agreement between two MSUs concurrently operating on different satellites gives an indication of instrument stability. Zonal means of the root-mean-square (r.m.s.) differences between anomalies from different satellites for two overlap periods (Fig. 3) reveal that retrievals from MSU-2R contain greater noise than those from MSU-2; this is because of the magnification of small differences between the relatively large radiances from multiangle views. MSU-2R retrievals also lack limb correction and retain fewer observations, so that the reproducibility of brightness temperatures between different satellites is not as good.

One source of this greater noise in MSU-2R arises from non-oxygen emission through the influence of surface emissivity. MSU-2R (MSU-2) brightness temperatures receive roughly 80% (90%) of their full signal from the atmosphere and 20% (10%) from surface emissions over land at sea level, while 90% (95%) of the signal is atmospheric over oceans¹¹. Significant variations in surface emis-

sion arise from changes in surface skin temperature, wetness, snow cover, and vegetation and thereby affect measures of interannual variability and trends: this noise is not random because tropical soil moisture, for instance, is affected by ENSO. Soil wetness can reduce a dry land microwave surface emissivity by 20–50% locally making the land appear colder. For a realistic change in surface emissivity of -0.2 at an unvegetated tropical land site, the MSU-2R (MSU-2) anomaly is $\sim -1.6^\circ\text{C}$ (-0.8°C). The sensitivity to a 1°C change in tropical land skin temperature is 0.2°C (0.1°C) for MSU-2R (MSU-2)¹¹. Diurnal variations in skin temperature over deserts and mountains are ~ 10 – 20°C , which would add 2 – 4°C of daily variability to MSU-2R. All of these aspects are consistent with the much greater r.m.s. differences between satellite records over land⁵. But because surface emissivity influences on the MSU-2R record are obfuscated by the procedures used for removing the diurnal cycle³, it is not possible to deconvolve the surface effects.

We estimate from the MSU-2R (MSU-2) data—using empirical orthogonal function analysis of spatial fields—that there are roughly eight (five) independent estimates of tropical tropospheric temperatures for both the atmosphere and the model. Monthly mean area-averaged tropical MSU-2R temperature differences between two satellites therefore have a standard error of $\sim 0.15^\circ\text{C}$ (Fig. 3) divided by $\sqrt{8}$, or $\sim 0.05^\circ\text{C}$, so that biases are known to $\sim \pm 0.1^\circ\text{C}$, which is large enough to corrupt long-term trends when several satellite records are merged, especially when the overlap is short as is often the case. Therefore, the cumulative evidence strongly suggests the recent coldness and the overall downward trend in MSU-2R temperatures are spurious and arise from difficulties in matching records between satellites combined with surface emissivity influences. The latter add considerable noise, especially over land, while the merging of satellite records requires long overlaps between different satellites and stable orbits that are not always achievable.

The magnitude of the MSU-2R discrepancy with tropical SSTs before late 1981 ($\sim 0.25^\circ\text{C}$) is larger than can be explained at present. Moreover, it is global and is reflected in MSU-2 anomalies although with much smaller magnitude (Fig. 1). J. R. Christy (personal communication) cites good agreement with radiosonde temperature records from 96 (mostly extratropical) sites during these years. Over the tropics, however, there are few reliable upper-air records to help resolve the differences. Before mid-1981 when data from NOAA-7 (a satellite with an afternoon Equator-crossing time) became available, the MSU record came from TIROS-N (in 1979) and NOAA-6 (a morning satellite) and the diurnal cycle was not well sampled. Tropical SST anomalies during this period were only slightly above the 1951–80 mean and are physically consistent with an index of the Southern Oscillation based on surface pressures, so the much warmer satellite temperatures cannot be accounted for. The second transition of concern is between the morning satellites NOAA-10 and NOAA-12 in mid-1991. Both could be matched to the afternoon satellite NOAA-11, but a drift in the Equator-crossing time of the NOAA-11 orbit of >3 hours degrades the matching process³. Moreover the eruption of Mt Pinatubo was just before the transition from NOAA-10 to NOAA-12, perhaps further confounding that match-up.

Statistical model fitting¹⁹ using SSTs objectively identifies two step-function break-points in the MSU-2R record in mid-1981 and mid-1991 which can be identified with the NOAA-6 to NOAA-7 and NOAA-10 to NOAA-12 transitions. With offsets of 0.23°C before July 1981 and -0.12°C after August 1991 included as adjustments to tropical MSU-2R, the monthly variance accounted for by linear regression (compare Fig. 1) rises from 61% to 80%; moreover, the residuals are small and are modelled statistically by a first-order autoregressive process, so the surface and satellite records can be completely reconciled within expected random noise levels.

Accordingly, the accumulated evidence indicates that there should be a small positive trend in MSU-2R, such as found for

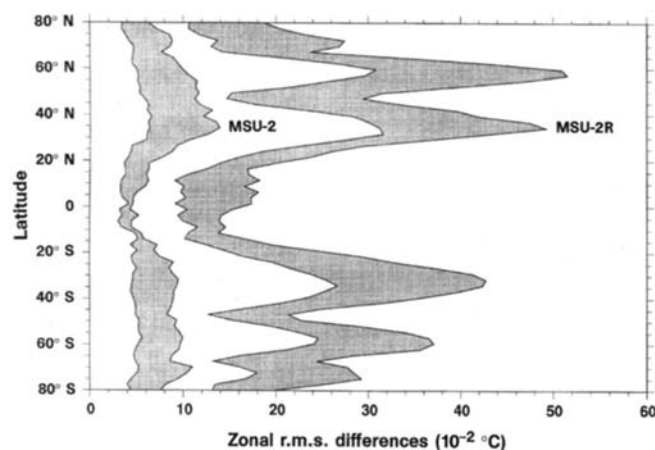


Figure 3 Zonal means of the r.m.s. differences in monthly MSU anomaly temperatures for the two overlap periods from satellites that differ most for MSU-2R (right) and MSU-2 (left). The NOAA-6 to NOAA-7 transition differences make up the minimum curves, and the match-up of NOAA-6 to NOAA-9 the maximum curves; other transitions fall in the shaded regions.

MSU-2; a trend also borne out by physical evidence of the retreat of tropical alpine glaciers and ice caps²⁰ and changes in freezing levels²¹. Although the MSU-2R trend has significant uncertainties and should be interpreted with caution, we conclude that the satellite record is otherwise excellent for examining interannual variability of tropospheric temperature. □

Received 29 October 1996; accepted 4 February 1997.

- Houghton, J. T. et al. (eds) *Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, 1996).
- Hansen, J., Ruedy, R., Sato, M. & Reynolds, R. *Geophys. Res. Lett.* **23**, 1665–1668 (1996).
- Christy, J. R., Spencer, R. W. & McNider, R. T. *J. Clim.* **8**, 888–896 (1995).
- Spencer, R. W., Christy, J. R. & Grody, N. C. *J. Clim.* **3**, 1111–1128 (1990).
- Spencer, R. W. & Christy, J. R. *J. Clim.* **5**, 858–866 (1992).
- Economist* 23 March, 83–85 (1996).
- Michaels, P. J. *Forging Consensus: Climate Change and the United Nations* (George C. Marshall Inst., Washington DC, 1996).
- Michaels, P. J. (ed.) *World Climate Report Vol. 1*, No. 17 (1996).
- Singer, S. F. *Science* **271**, 581–582 (1996).
- Spencer, R. W. in *State of the Climate Report A World in Perspective* (ed. Michaels, P. J.) 9–11 (Western Fuels Assoc. Arlington, VA, 1996).
- Shah, K. P. & Rind, D. *J. Geophys. Res.* **100**, 13841–13874 (1995).
- Hurrell, J. W. & Trenberth, K. E. *J. Clim.* **9**, 2222–2232 (1996).
- Hansen, J. et al. *Clim. Change* **30**, 103–117 (1995).
- Christy, J. R. & McNider, R. T. *Nature* **367**, 325 (1994).
- Jones, P. D. *Geophys. Res. Lett.* **21**, 1149–1152 (1994).
- Smith, T. M., Reynolds, R. W., Livezey, R. E. & Stokes, D. C. *J. Clim.* **9**, 1403–1420 (1996).
- Reynolds, R. W. & Smith, T. M. *J. Clim.* **7**, 929–948 (1994).
- Kiehl, J. T. et al. *NCAR Tech. Note NCAR/TN-420+STR* (National Center for Atmospheric Research, Boulder, CO, 1996).
- Jones, R. H. & Dey, I. *Chem. Phys. Lipids* **76**, 1–6 (1995).
- Thompson, L. G. et al. *Science* **269**, 46–50 (1995).
- Diaz, H. F. & Graham, N. E. *Nature* **383**, 152–155 (1996).

Acknowledgements. We thank J. Christy for the data for Fig. 3, and B. Bailey for assistance with the statistical analysis. The National Center for Atmospheric Research is sponsored by the US NSF. This work was supported in part by the US National Oceanic and Atmospheric Administration's Climate and Global Change Program, and NASA.

Correspondence should be addressed to J.W.H. (e-mail: jhurrell@ucar.edu).

Coding of intention in the posterior parietal cortex

L. H. Snyder, A. P. Batista & R. A. Andersen

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

To look at or reach for what we see, spatial information from the visual system must be transformed into a motor plan. The posterior parietal cortex (PPC) is well placed to perform this function, because it lies between visual areas, which encode spatial information^{1,2}, and motor cortical areas. The PPC contains several subdivisions, which are generally conceived as high-order sensory areas^{3,4}. Neurons in area 7a and the lateral intraparietal area fire before and during visually guided saccades. Other neurons in areas 7a and 5 are active before and during visually guided arm movements^{5–10}. These areas are also active during memory tasks in which the animal remembers the location of a target for hundreds of milliseconds before making an eye or arm movement. Such activity could reflect either visual attention^{11–15} or the intention to make movements^{16–22}. This question is difficult to resolve, because even if the animal maintains fixation while directing attention to a peripheral location, the observed neuronal activity could reflect movements that are planned but not executed²². To address this, we recorded from the PPC while monkeys planned either reaches or saccades to a single remembered location. We now report that, for most neurons, activity before the movement depended on the type of movement being planned. We conclude that PPC contains signals related to what the animal intends to do.

Neurons in the PPC were recorded from three hemispheres of two adult macaque monkeys during interleaved delayed-saccade and delayed-reach trials (Fig. 1). Delay activity (measured 150–600 ms

after target extinction) was significantly modulated by direction of movement during either or both tasks in 373 of 652 neurons for which complete data was collected (Student's *t*-test, $P < 0.05$). Of these, 68% were motor-intention specific: 21% were significantly modulated before eye but not arm movements, while 47% were significantly modulated before arm but not eye movements. Surprisingly, activity during the cue interval (50 ms before, to 150 ms after, extinction) was intention specific in 44% of the 443 active neurons. Specificity so early after target presentation suggests that these findings apply during saccades and reaches made without delays, that is, during saccades and reaches to visible targets.

A dissociation task was introduced to control for the possibility that the animal planned both a reach and a saccade to a target even when only a single movement was instructed (Fig. 2). In fact, in a previous pair of studies^{12,23} which reported no specificity for saccades compared to reaching movements in the PPC, animals trained to reach towards targets without looking at them nonetheless looked towards the target at the end of the trial. It is likely that plans for both eye and arm movements to the target were formed simultaneously, with the execution of the eye movement delayed until the end of the trial. Delayed and even entirely unexecuted plans for movement may influence firing in the lateral intraparietal (LIP) area²². Similarly, in a delayed "go/no-go" task, neurons in area 5 code target location regardless of whether or not a

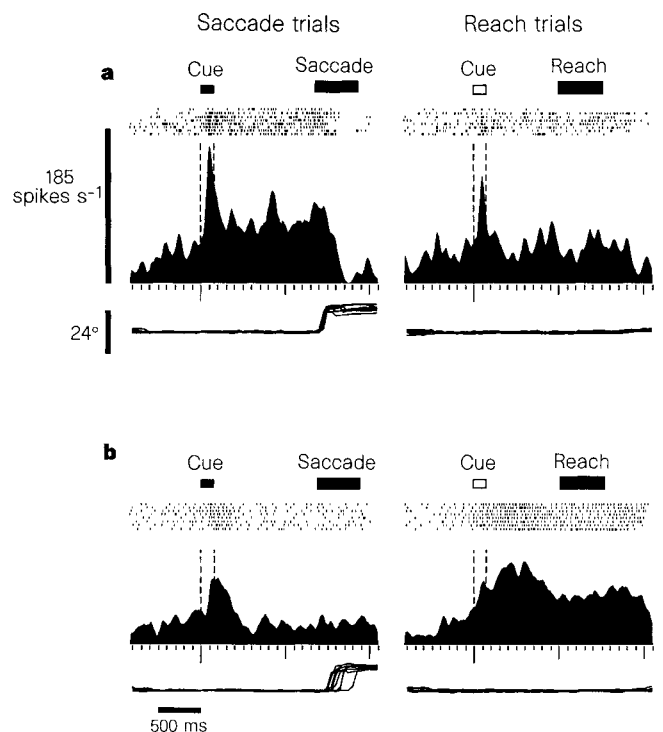


Figure 1 Responses of two intention-specific neurons in the delayed-saccade (left) and delayed-reach (right) tasks. Each panel shows timing of peripheral flash ('Cue': red flashes indicated by filled bars, green flashes by open bars) and response ('Saccade' or 'Reach'); eight rows of rasters corresponding to every third action potential recorded during each of eight trials; a spike density histogram of neuronal activity, generated by convolution with a triangular kernel²⁷ aligned on cue presentation, with cue onset and offset indicated by dashed lines; and eight overlaid traces showing vertical eye position. Neuronal responses in the cue interval (50 ms before to 150 ms after cue offset) were nonspecific. However, during the delay interval (150–600 ms), firing depended specifically on motor intent. **a**, A cell showing elevated delay period firing before a saccade (left) but not before a reach (right). For illustration purposes, data for this cell were collected using a fixed delay interval. **b**, A second cell showing reach rather than saccade specificity during the delay interval.

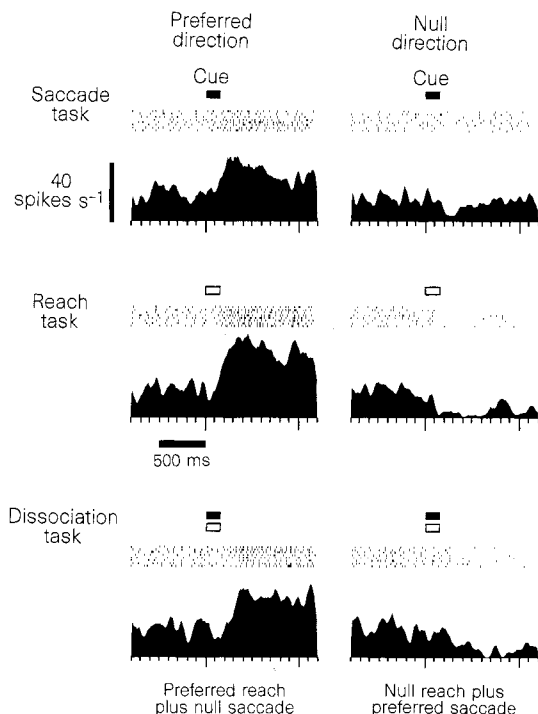


Figure 2 An intention-specific neuron whose motor specificity was revealed by the dissociation task. Delay activity was greater before movements towards the receptive field ('Preferred direction', left column) compared to away ('Null direction', right column) in both delayed saccade (top row) and reach (middle row) tasks. Thus in single-movement tasks, the neuron appears to code remembered target location independent of motor intent. However, motor specificity was revealed in the dissociation task (bottom row). Firing was vigorous before a preferred reach combined with a null saccade (bottom left), but nearly absent before a preferred saccade plus null reach (bottom right). Thus when both a reach and a saccade were planned, delay activity reflected the intended reach and not the intended saccade. Panel formats are similar to Fig. 1. Every other action potential is indicated by one raster mark.

movement is made²⁴. The dissociation task eliminates plans for movements that will not be executed by explicitly instructing eye and arm movements in opposite directions. Of neurons with non-specific delay activity in the single-movement tasks (delayed reach or delayed saccade) that were tested in the dissociation task, 62% were revealed to be intention specific, bringing the total percentage of specific neurons to 84% (Table 1). In the cue interval, corresponding percentages were 45% and 63%.

Neurons specific for eye and arm movements were anatomically segregated (Table 1 and Fig. 3). While cells throughout the PPC showed motor-specific responses, cells in two subregions (area LIP and a reach area medial and posterior to LIP) tended to have strong, prolonged delay activity. In the middle third of the longitudinal extent of the intraparietal sulcus, intended-eye-movement cells outnumbered intended-arm-movement cells by 5:1. Of 47 cells active during the delay period, 28 were eye specific and only 5 arm specific in the simple tasks, with an additional 4 eye-specific and only 1 arm-specific cell revealed by the dissociation task. In a second area, medial and posterior to LIP, arm-specific cells outnumbered eye-specific cells 9:1. Of 95 active cells, 68 were arm specific and only 9 eye specific in the simple tasks, with an additional 12 arm-specific cells revealed by the dissociation task. This anatomical segregation argues against chromatic tuning as a basis for our results, as clustering of red- or green-preferring neurons over many square millimetres of cortex is unlikely. The dissociation task revealed neurons specific for saccades as well as reaches. This helps to rule out the possibility that different activity levels reflect differential allocations of attention; if reaching, for example, required greater attention, then all cells should have appeared reach specific.

Our results reveal separate intended-reach and intended-saccade pathways in the PPC. This demonstrates that the decision of how to utilize a particular sensory stimulus is reflected in PPC firing. Attention may still be present in the PPC^{11,12,14,25} encoded by the small number of cells that are not specific for one type of movement, or in the nonspecific cue responses of the cells that are movement-specific in the delay period, or in the weak response of some specific neurons before their non-preferred movement. Alternatively, nonspecific neurons may reflect plans for moving body parts other than the eyes or arms (for example, pinna movements) that

Table 1 Summary of neurons

	All areas	Area LIP		Reach area	
	Both	M1	M2	M1	M2
Cue interval (100–300 ms from cue onset)					
Saccade specific	104 (23%)	17 (40)	12 (39)	17 (24)	1 (4)
Plus dissociation	161 (36%)	26 (62)	22 (71)	18 (26)	1 (4)
Reach specific	91 (21%)	4 (10)	1 (3)	13 (19)	13 (50)
Plus dissociation	119 (27%)	4 (10)	1 (3)	17 (24)	17 (65)
Nonspecific	163 (37%)	12 (29)	8 (26)	35 (50)	8 (31)
Total active cells	443	42	31	70	26
Delay interval (300–750 ms from cue onset = 150–600 ms from cue offset)					
Saccade specific	79 (21%)	17 (59)	11 (61)	9 (13)	0 (0)
Plus dissociation	87 (23%)	18 (62)	14 (78)	9 (13)	0 (0)
Reach specific	175 (47%)	3 (10)	2 (11)	47 (68)	21 (81)
Plus dissociation	227 (61%)	4 (14)	2 (11)	55 (80)	25 (96)
Nonspecific	59 (16%)	7 (24)	2 (11)	5 (7)	1 (4)
Total active cells	373	29	18	69	26

Most active cells were intention specific in both cue and delay intervals. Data for the cue interval (during and immediately after stimulus presentation) and for the delay interval (after stimulus presentation but well before movement) are shown. Columns show cell counts and percentages for cells in all areas, LIP cells in the first and second monkey (M1 and M2), and reach-area cells in M1 and M2. Effects of movement intention were assayed by comparing activity before movements in the best and opposite directions (determined in a previous block of trials). If activity was modulated only by the intention to saccade in opposite directions (Student's two-tailed *t*-test, $P < 0.05$), the cell was classified as saccade specific. If activity was modulated only by the intention to reach in opposite directions, the cell was classified as reach specific. If activity was modulated by either movement, the cell was classified as nonspecific. Rows labelled 'plus dissociation' include cells active in both the simple saccade and reach tasks whose firing depended on movement direction in the dissociation task. Inactive cells and cells without directionally selective activity were excluded. Not all cells were tested in the dissociation task.

we did not test²⁶. Some saccade-specific neurons are active before arm movements and some reach-specific neurons are active before saccades. We propose that this activity reflects plans for movements not explicitly called for by the task, but formed automatically in response to target appearance. When these default plans are countermanded by explicit instructions, as in the dissociation task, intention specificity is revealed. The mechanism may involve inhibition between neurons in the same motor pathway coding different directions. Functionally, this coupling of saccade and reach activity may reflect the fact that these movements are often coupled.

The idea that the PPC plays a role in motor planning is consistent with previous experiments in area LIP. In a delayed double-saccade task, animals memorized two flashed locations and then, after a delay, saccaded to them sequentially. Most LIP delay activity coded the goal of the first saccade (target 1) rather than the location of the most recently presented stimulus (target 2). At the time of the first saccade, firing changed to code the goal of the second saccade (target 2). These two observations, taken together, rule out a strictly sensory role for area LIP, and support the motor planning hypothesis^{20–22}. The appearance of activity coding target 2 after the first saccade can alternatively be explained as sensory remapping of a remembered stimulus in retinal coordinates. The observation that this activity sometimes anticipates the completion of the first saccade has been proposed to support the sensory remapping hypothesis¹³. However, predictive behaviour is as likely to occur in motor planning as in sensory pathways, and so the observation of

anticipatory activity does not favour either hypothesis. In contrast, we now have shown directly that the majority of the delay activity in LIP as well as in the neighbouring reach area is related to specific motor intention and not to either sensory stimuli or spatial attention.

Other instances of task requirements influencing PPC responses have been reported, and have been ascribed to attentional processes^{11,12}. In the dorsal visual stream, posited to be dedicated to action^{2,6}, attention and intention can be difficult to distinguish. Even when a given task does not require an action, plans for movements not explicitly required by the task can nonetheless be formed, producing potentially deceptive results. The results reported here indicate that it is important to rule out intention-related signals (using controls like the dissociation task) before concluding that task-dependent modulation in the PPC reflects an attentional process. □

Methods

Animals faced an array of nine buttons 3.7 cm in diameter at 28 cm distance. Each button contained a red and a green light-emitting diode (LED) side by side behind a 1.2-cm translucent lens. The animal pressed buttons illuminated green using its contralateral arm, and fixated buttons illuminated red. No other lights were present. All trials began with fixation ($\pm 2.7^\circ$) and depression of the illuminated central button. After 750 ms, a red (saccade task) or green (reach task) peripheral LED was flashed for 150 ms. After a 1–1.6 s delay, the central LEDs were extinguished and the monkey saccaded (latency mean $\pm$ s.d.,

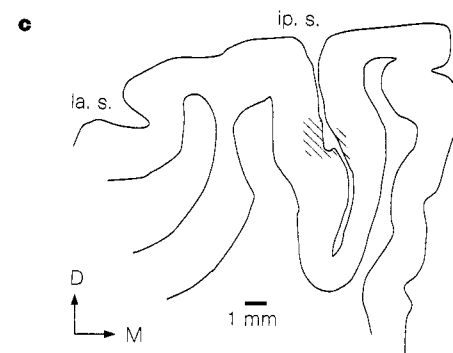
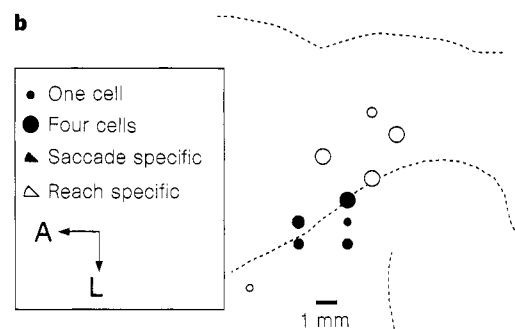
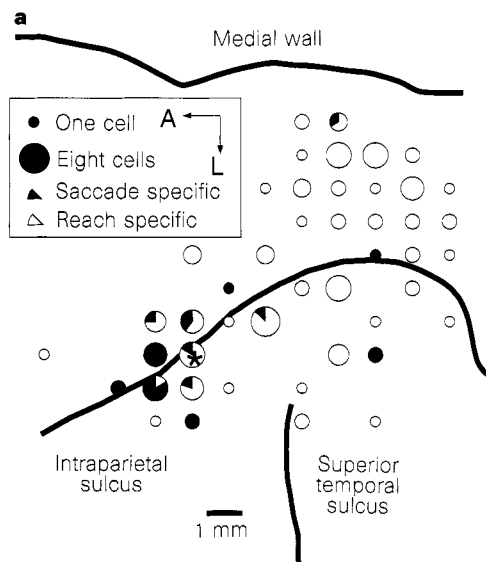


Figure 3 Neurons specific for the intention to make eye and arm movements were anatomically segregated. **a**, Surface map showing the locations of motor-specific cells relative to the intraparietal and superior temporal sulci in the left hemisphere of one animal. The area of each pie chart corresponds to the number of cells (between one and eight) with saccade-specific (filled) and reach-specific (open) delay activity at each location. Only cells with robust delay activity are shown (≥ 16 spikes per s in one or both task delay intervals). The excluded cells are primarily located in and around area 7a, the ventral intraparietal area and the medial superior temporal area. Because the x-y positioning apparatus for single neuron recording and for dye injection were not identical, the placement of tracks may be misaligned by up to ± 1 mm. **b**, Same format for robust motor-specific

cells from the right hemisphere of the second animal. Anatomy is not available, so the sulcal pattern of the first animal (dotted lines) is superimposed to indicate roughly where these neurons may lie. Regardless of how different the sulcal patterns in the two animals may be, the clustering of reach and saccade cells is quite similar. The sulcal pattern and recording sites have been reflected onto the left hemisphere to facilitate comparisons with **a**. **c**, Coronal section showing a nuclear yellow dye injection made into the centre of the cluster of eye-specific cells in the first animal (indicated by an asterisk in **a**). The injection (hatching in **c**) was visualized midway down the lateral bank of the intraparietal sulcus (ip.s.), in area LIP (la.s., lateral sulcus). Arrows indicate anterior, lateral, dorsal and medial directions (A, L, D, M).

172 $\pm$ 36 ms) or reached (269 $\pm$ 45 ms) to the remembered peripheral location in complete darkness. The animal maintained central fixation during reach trials, and maintained central button depression during saccade trials. Eight delayed-saccade and eight delayed-reach trials were performed in each of eight directions. For most neurons, the best direction was determined from this first block of trials and then a second block of interleaved delayed-saccade, delayed-reach and delayed-dissociation movements was performed in the best and the opposite directions (8 or 16 trials per task per direction). The dissociation task was similar to the simple tasks, but now simultaneous red and green flashes were delivered on opposite sides of the fixation point, and the animal responded with a near-simultaneous reach and saccade in opposite directions when the fixation light was extinguished. The animal typically performed over 90% of the trials successfully. Results (Table 1) were based primarily on data from the second block of trials.

Received 10 December 1996; accepted 14 January 1997.

1. Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (eds Ingle, D. J., Goodale, M. G. & Mansfield, R. J. W.) 549–586 (MIT Press, Cambridge, MA, 1982).
2. Goodale, M. G. & Milner, A. D. *Trends Neurosci.* 15, 220–225 (1992).
3. Hyvärinen, J. & Poranen, A. *Brain* 97, 673–692 (1974).
4. Gross, C. G. *Cereb. Cortex* 5, 455–469 (1994).
5. Hartje, W. & Eitlinger, G. *Cortex* 9, 346–354 (1974).
6. Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H. & Acuna, C. *J. Neurophysiol.* 38, 871–908 (1975).
7. Seal, J. & Commenges D. *Exp. Brain Res.* 58, 144–153 (1985).
8. Shibutani, H., Sakata, H. & Hyvärinen, J. *Exp. Brain Res.* 55, 1–88 (1984).
9. Andersen, R. A., Essick, G. K. & Siegel, R. M. *Exp. Brain Res.* 67, 316–322 (1987).
10. Murata, A., Gallese, V., Kaseda, M. & Sakata, H. *J. Neurophysiol.* 75, 2180–2186 (1996).
11. Robinson, D. L., Goldberg, M. E. & Stanton, G. B. *J. Neurophysiol.* 41, 910–932 (1978).
12. Bushnell, M. C., Goldberg, M. E. & Robinson, D. L. *J. Neurophysiol.* 46, 755–772 (1981).
13. Duhamel, J.-R., Colby, C. & Goldberg, M. E. *Science* 255, 90–92 (1992).
14. Steinmetz, M. A., Conner, C. E., Constantinidis, C. & McLaughlin, J. R. *J. Neurophysiol.* 72, 1020–1023 (1994).
15. Colby, C. L., Duhamel, J. R. & Goldberg, M. E. *Cereb. Cortex* 5, 470–481 (1995).
16. Gnadt, J. W. & Andersen, R. A. *Exp. Brain Res.* 70, 216–220 (1987).
17. Andersen, R. A. *Annu. Rev. Neurosci.* 12, 377–403 (1989).
18. Thier, P. & Andersen, R. A. *Proc. Natl Acad. Sci. USA* 93, 4962–4967 (1996).
19. Shadlen, M. N. & Newsome, W. T. *Proc. Natl Acad. Sci. USA* 93, 628–633 (1996).
20. Mazzoni, P., Bracewell, R. W., Barash, S. & Andersen, R. A. *J. Neurophysiol.* 76, 1439–1456 (1996).
21. Andersen, R. A. *Cereb. Cortex* 5, 457–469 (1995).
22. Bracewell, R. W., Mazzoni, P., Barash, S. & Andersen, R. A. *J. Neurophysiol.* 76, 1457–1464 (1996).
23. Goldberg, M. E. & Bushnell, M. C. *J. Neurophysiol.* 46, 773–787 (1981).
24. Kalaska, J. F. & Crammond, D. J. *Cereb. Cortex* 5, 410–428 (1995).
25. Lynch, J. C., Mountcastle, V. B., Talbot, W. H. & Yin, T. C. T. *J. Neurophysiol.* 40, 362–389 (1977).
26. Bon, L. & Lucchetti, C. *Exp. Brain Res.* 102, 259–271 (1994).
27. Scott, D. W. *Ann. Statist.* 13, 1024–1040 (1985).

Correspondence and requests for materials should be addressed to R.A.A. (e-mail: andersen@vis.caltech.edu).

Acknowledgements. We thank B. Gillikin for technical assistance and S. Gertmenian for editorial assistance. This work was supported by the National Eye Institute, Sloan Center for Theoretical Neurobiology at Caltech, Della Martin Foundation, and the Office of Naval Research.

Self-centring activity of cytoplasm

Vladimir I. Rodionov & Gary G. Borisov

Laboratory of Molecular Biology, University of Wisconsin, Madison, Wisconsin 53706, USA

Fish melanophore cells aggregate pigment granules at the centre or redisperse them throughout the cytoplasm. The granules move along radial microtubules by means of molecular motors^{1–3}. Cytoplasmic fragments of melanophores organize a radial array of microtubules and aggregate pigment at its centre^{4–7}. Here we report self-centring in microsurgically produced cytoplasmic fragments of black tetra melanophores. We observed rapid (10 min) formation of a radial microtubule array after stimulation of aggregation. Arrangement of microtubules in the fragments returned to random during pigment redispersion. Apparently, formation of the radial array does not depend on a pre-existing microtubule-organizing centre. The array did not form in granule-free fragments nor in fragments treated with inhibitors of the intracellular motor protein cytoplasmic dynein. We conclude that

formation of the radial microtubule array is induced by directional motion of pigment granules along microtubules and present evidence that its position is defined by interaction of microtubules with the surface.

Nascent fragments of cultured black tetra melanophores, which had been induced to aggregate, initially formed the aggregate at the cut edge. Remarkably, after 2–3 min delay, the aggregate relocated to the centre (Fig. 1a) with kinetics approximated by a single declining exponential ($t_{1/2} = 150 \pm 55$ s, mean $\pm$ s.d.; $n = 10$). Fragments induced to aggregate after a period of incubation (30–300 min) showed irregular granule motion which resulted in aggregation at the centre (Fig. 1a). Immunostaining with anti-tubulin antibody confirmed that nascent fragments, like intact cells, displayed an array of microtubules directed towards the parental cell centre, whereas in incubated fragments microtubules were randomly arranged (Fig. 1a). Remarkably, upon inducing aggregation, both nascent and incubated fragments formed a radial array of microtubules focusing on the granule aggregate situated at the fragment centre (Fig. 1a). Fragments responded to caffeine treatment by dispersing pigment granules indistinguishably from intact cells. As dispersion depends on a microtubule plus-end kinesin-like motor⁸, we may infer that the polarity of microtubules in the radial array was also similar to that of intact cells: minus ends at the aggregate and plus ends towards the periphery. The microtubule-stabilizing drug taxol did not interfere with pigment aggregation in intact cells, but blocked redistribution of the pigment aggregate to the centre in nascent fragments (Fig. 1a), and fusion and centring of local clusters in incubated fragments (Fig. 1a); it also prevented formation of a radial array of microtubules (not shown). Thus, fragments of black tetra melanophores were able to form and centrally locate a radial array of microtubules. Centring required microtubule dynamics and, contrary to previous observations^{5–7}, was very rapid.

To test whether the radial microtubule array in the fragments was organized by a centrosome, we tested for molecular components of microtubule-nucleating structures, γ -tubulin⁹ and pericentrin¹⁰. Double immunostaining showed that γ -tubulin was at the focus of the microtubule array in parental cells (Fig. 1b), but could not be detected in the fragments even after incubation (300 min) (Fig. 1b) or after a cycle of aggregation–redispersion. Immunostaining for pericentrin was positive in the parental controls but negative in the cytoplasmic fragments (result not shown). Thus, we found no evidence that the fragments either retained or reconstituted a centrosome after severing from their parent cell.

Direct live observations of labelled microtubules showed that the random distribution of microtubules in incubated fragments transformed into a radial array during aggregation and returned back to a random arrangement during the course of redispersion (not shown). Thus, microtubule distribution seems to be coupled to the aggregation state of pigment granules. To test whether the attainment of a radial microtubule array depends on the interaction of microtubules with granules, we prepared fragments devoid of pigment granules by dissecting them from the cells with aggregated pigment. Adrenaline treatment of pigment-free fragments failed to induce microtubule reorganization (Fig. 1c), suggesting that formation of the radial array depends on the motion of granules along microtubules.

To test further the role of granule motion in microtubule reorganization, we used inhibitors of cytoplasmic dynein and kinesin, microtubule-dependent motors responsible for pigment aggregation and dispersion^{2,3,8}. The cytoplasmic dynein inhibitor sodium orthovanadate^{11,12}, which prevents pigment aggregation in intact chromatophores^{13,14}, blocked pigment aggregation and formation of radial microtubules in the fragments (Fig. 2). Other dynein inhibitors erythro-9-[3-(2-hydroxy-nonyl)] adenine^{11,12} (EHNA; 2 mM) and AMP-PNP^{11,12} (100 mM in the needle), though not entirely specific for pigment aggregation, had the

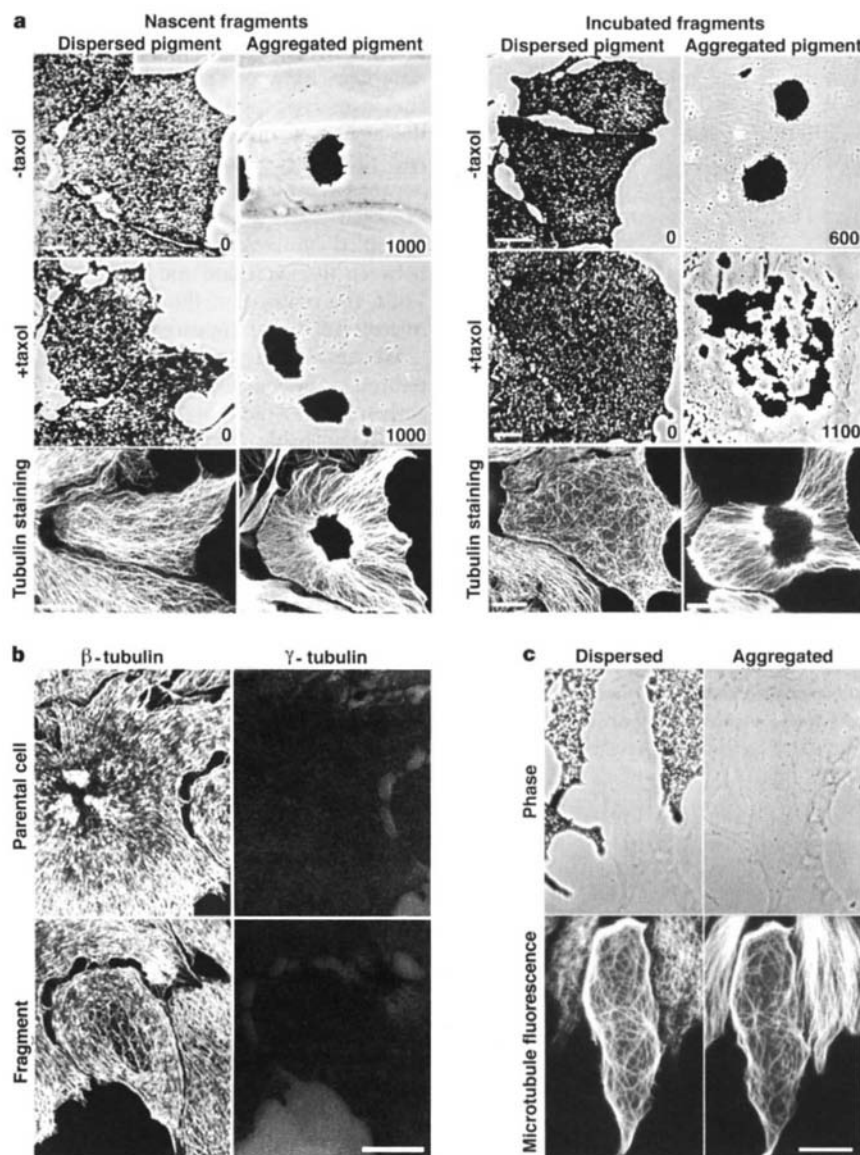


Figure 1 Formation of radial microtubule (MT) array in melanophore fragments does not depend on MT nucleation at a new centre and requires pigment granules. **a**, Patterns of pigment motion and MT distribution in nascent fragments (left) and in fragments incubated for 120 min after dissection (right); pigment aggregation in fragments dissected from non-treated cells (upper panels) and from cells pretreated with taxol (10^{-5} M) for 15 min (middle panels); MT distribution in nascent fragments with dispersed (lower panels, left-hand micrographs) and aggregated (lower panels, right micrographs) pigment. Numbers indicate time in seconds after stimulation with adrenaline. Scale bars, $20\ \mu\text{m}$. **b**, Immu-

nostaining for γ -tubulin of intact cells and fragments; immunofluorescence micrographs showing distribution of MTs (left) and γ -tubulin (right) for intact cell (top) and fragment incubated for 180 min after dissection (bottom); scale bar, $40\ \mu\text{m}$; **c**, Microtubules in a fragment lacking pigment granules. Phase contrast images (top) and live fluorescence images of MTs (bottom) before (left) and after (right) adrenaline treatment. Reorganization of MTs did not occur in the absence of granules. Parental cell that served as a control responded normally to caffeine and adrenaline by pigment redispersion and aggregation. Scale bars, $20\ \mu\text{m}$.

same effect ($n = 10$ for each). In contrast, selective inhibition of kinesin-like proteins with antibody HD⁸ (Fig. 2, bottom row) and AMP-PNP¹⁵ at low concentration (10 mM in needle; $n = 10$) did not prevent aggregation and organization of radial microtubules, but inhibited pigment dispersion and stabilized microtubules in the radial arrangement (Fig. 2). Control experiments showed that inhibitors of intracellular motors did not suppress microtubule dynamics. These results indicate that formation and dissociation of the radial array depends on motion of pigment granules along microtubules that is mediated by molecular motors.

How does the pigment aggregate that marks the focal point of radiating microtubules become positioned at the centre? We considered that the centring mechanism involved the interaction of microtubules with the fragment surface and tested this relationship

by varying the geometry of the fragment. As the general pattern of aggregation was to the centroid of a fragment, we devised a torus shape, in which the centroid lay outside the fragment. After stimulation with adrenaline, granules in a toroidal fragment moved to a zone equidistant from both the outer (pre-existing) and inner (newly formed) edges of the fragment, whereas granules in the discoidal remnant of the parental cell aggregated to the centroid as normal (Fig. 3). Thus, consistent with a surface-determined mechanism, granules in the toroid moved away from the margins, even though the direction for about half of them was opposite to that for granules in the intact cell or remnant discoid.

During the centring motion, microtubules extending from the pigment mass to the margin must elongate. To test whether new subunits are added at the distal ends and microtubules themselves

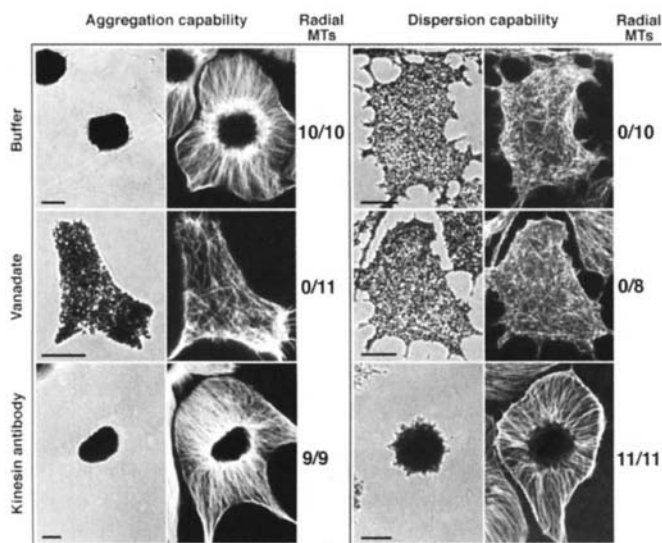


Figure 2 Inhibitors of motor activity interfere with microtubule (MT) rearrangement in melanophore fragments. Pigment distribution (left columns) and arrangement of MTs (right columns) in melanophore fragments injected with buffer (top row) sodium orthovanadate (middle) or kinesin antibody (bottom row) and stimulated to aggregate (left composite panel) or disperse (right composite panel) pigment granules; numbers at side of each panel indicate the number of fragments with radial arrangement of MTs per total number of fragments tested. Scale bars, 20 μ m.

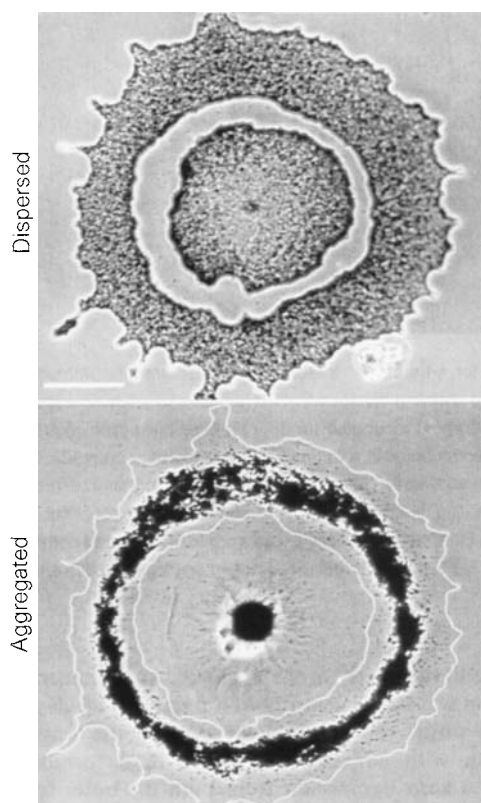


Figure 3 Pigment motion in a toroidal-shaped fragment. A melanophore was dissected in a ring path with a microneedle producing a toroidal cytoplasmic fragment and a discoidal remnant containing the nucleus and centrosome (top). Pigment aggregation was induced with adrenaline (bottom). In both cases pigment in the toroidal fragment moved away from the fragment surface, forming an annular ring, whereas pigment in the discoidal fragment aggregated centripetally as normal. Scale bar, 50 μ m.

are transported during the centring motion, we labelled microtubules in the nascent fragment with rhodamine-tubulin and used photobleaching to create reference marks on microtubules in the course of redistribution of pigment aggregate to the centre. As the aggregate moved, the distance between the proximal edge and the bleached zone remained unchanged while the distance between the bleached zone and the aggregate increased (Fig. 4). Bleached zones placed on the distal side of the aggregate also remained stationary relative to the edge while the distance between the aggregate and the bleached zone decreased (Fig. 4). Thus, the centring of the aggregate occurred on a population of microtubules that appeared to be stationary.

We have shown here that cytoplasmic fragments of melanophores, in the course of pigment aggregation, organize and position at their centre a radial array of microtubules with functional polarity indistinguishable from that of intact cells. On the basis of our results, we propose a model for the self-centring of melanophore cytoplasm. We propose that adrenaline treatment triggers dynein-like motors to translocate pigment granules to the minus ends of microtubules. As each granule can presumably interact with more than one microtubule, this motion brings microtubule minus ends together and thus induces formation of a radial array of microtubules. Such a mechanism has been proposed to explain the organization of microtubule asters¹⁶ and spindle poles¹⁷ in the presence of cytoplasmic dynein in mitotic *Xenopus* egg extracts, as well as microtubule asters of reversed polarity in the presence of

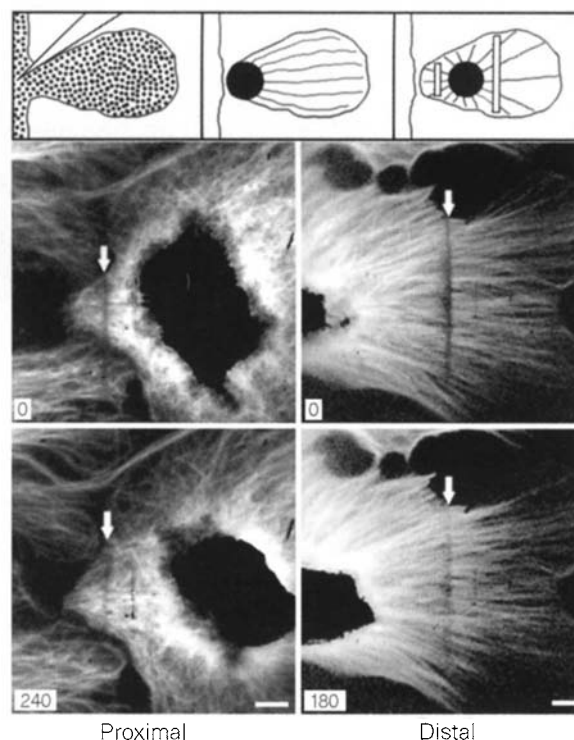


Figure 4 Photobleaching of MTs in the course of centripetal motion of pigment aggregate in nascent fragment. Top, scheme of the experiment. Intact melanophores were injected with rhodamine-labelled tubulin subunits and 60 min later fragments were dissected and stimulated with adrenaline. When pigment aggregate formed and started to move to the centre, MTs were photobleached with a laser microbeam focused into a narrow zone and positioned perpendicular to the path of motion of the aggregate and between the aggregate and either the fragment's proximal (cut) or distal edge. Images of the bleached zone were obtained immediately after irradiation (middle) and a few minutes later when the aggregate shifted to the fragment's centre (bottom). Left-hand photos show results of photobleaching between the aggregate and the proximal edge; right-hand photos, between the aggregate and the distal edge. Numbers indicate time in seconds after irradiation. Scale bars, 5 μ m.

kinesin and chromaffin granules *in vitro*¹⁸. The information for locating the centre is derived from the geometrical relationship of the aggregate to the surface. The driving force for centring is presumably provided by motion of pigment granules coupled to the addition of new subunits to the microtubule minus ends, although contribution of dynamics at the microtubule plus ends is still a possibility. We believe that the self-organization mechanism described here is of general importance and contributes to formation and positioning of microtubule radial arrays in non-pigment cells as well, although the nature of motors and organelles involved remains to be determined. Thus, the mystery of self-centring can be explained simply in terms of the known properties of microtubules: their dynamic behaviour and interaction with motor proteins operating under the geometrical constraint of the surface. □

Methods

Melanophore cultures. Tissue cultures of black tetra melanophores were prepared as described¹⁹. Aggregation of melanosomes was triggered with 10^{-5} M adrenaline and pigment redispersion was induced by 5 mM caffeine.

Microsurgery and microinjection. To prepare fragments, melanophore processes were dissected with microneedles of 0.1 μ m tip diameter. To obtain fragments free of pigment granules, intact cells were first induced to aggregate their pigment and a fragment was then dissected from the peripheral cytoplasm. Microinjections were done as described¹⁹. Tubulin labelled with X-rhodamine was prepared as described²⁰ and used at a needle concentration of 10–15 mg ml⁻¹. Fragments with labelled microtubules were obtained by injection of parental cells with X-rhodamine–tubulin and subsequent dissection of the fragments. Sodium orthovanadate and kinesin antibody HD were used at needle concentrations of 500 μ M and 15 mg ml⁻¹ respectively. After injection with antibody or orthovanadate, fragments were stimulated to aggregate or disperse pigment granules and 16 min later were fixed with glutaraldehyde and immunostained with anti- β -tubulin antibody.

Immunostaining. Immunostaining with β -tubulin antibody was done as before²⁰. For double immunostaining with antibodies specific for β - and γ -tubulin, intact cells and fragments were fixed with methanol at -20°C for 6 min, rehydrated in phosphate-buffered saline, and successively incubated with a rabbit monospecific antibody against a peptide corresponding to residues 38–53 of *Schizosaccharomyces pombe* γ -tubulin (gift from G. Hering), rhodamine-conjugated goat anti-rabbit antibody, mouse monoclonal anti- β -tubulin antibody (Amersham) and fluorescein-conjugated anti-mouse antibody.

Image acquisition. Phase-contrast images of fragments and intact cells were acquired with a video camera and recorded on an optical memory disk recorder as 16-frame averages in the time-lapse mode (one image per 10 s). Live fluorescent images with a charge-coupled device (CCD) camera were acquired as described²⁰.

Photobleaching. Photobleaching of microtubules labelled with X-rhodamine–tubulin was done as described²⁰.

Received 18 November 1996; accepted 20 January 1997.

- Schliwa, M. Mechanisms of intracellular organelle transport. in *Cell and Muscle Motility* (ed. Shay, J. W.) 5, 1–82 (Plenum, New York and London, 1984).
- Obika, M. Cytophysiology of fish chromatophores. *Zool. Sci. (Tokyo)* 3, 1–11 (1986).
- Haimo, L. H. & Thaler, C. D. Regulation of organelle transport: lessons from color change in fish. *BioEssays* 16, 727–732 (1994).
- Matthews, S. Observations on pigment migration within the fish melanophore. *J. Exp. Zool.* 58, 471–486 (1931).
- McNiven, M., Wang, M. & Porter, K. R. Microtubule polarity and the direction of pigment transport reverse simultaneously in surgically severed melanophore arms. *Cell* 37, 753–765 (1984).
- McNiven, M. & Porter, K. R. Microtubule polarity confers direction of pigment transport in chromatophores. *J. Cell Biol.* 103, 1547–1555 (1986).
- McNiven, M. & Porter, K. R. Organization of microtubules in centrosome-free cytoplasm. *J. Cell Biol.* 106, 1593–1605 (1988).
- Rodionov, V. I., Gyoeva, F. K. & Gelfand, V. I. Kinesin is responsible for centrifugal movement of pigment granules in melanophores. *Proc. Natl. Acad. Sci. USA* 88, 4956–4960 (1991).
- Oakley, B. γ -Tubulin. In: *Microtubules* (eds Hyams, J. S. and Lloyd, C. W.) 33–45 (Wiley-Liss, New York, 1994).
- Doxsey, S. J., Stein, P., Evans, L., Calarco, P. & Kirschner, M. Pericentriolar, a highly conserved centrosome protein involved in microtubule organization. *Cell* 76, 639–650 (1994).
- Paschal, B. M. & Vallee, R. B. Retrograde transport by the microtubule-associated protein MAP 1C. *Nature* 330, 181–183 (1987).
- Shpetner, H. S., Paschal, B. M. & Vallee, R. B. Characterization of microtubule-activated ATPase of brain cytoplasmic dynein (MAP 1C). *J. Cell Biol.* 107, 1001–1009.
- Beckerle, M. C. & Porter, K. R. Inhibitors of dynein activity block intracellular transport in erythrocytes. *Nature* 295, 701–703 (1982).

- Clark, T. G. & Rosenbaum, J. L. Pigment particle translocation in detergent-permeabilized melanophores of *Fundulus heteroclitus*. *Proc. Natl. Acad. Sci. USA* 79, 4655–4659 (1982).
- Vale, R. D., Reese, T. S. & Sheetz, M. P. Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. *Cell* 42, 39–50 (1985).
- Verde, F., Berrez, J.-M., Antony, C. & Karsenti, E. Taxol-induced microtubule asters in mitotic extracts of *Xenopus* eggs: requirement for phosphorylated factors and cytoplasmic dynein. *J. Cell Biol.* 112, 1177–1187 (1991).
- Heald, R., Tournebise, R., Blank, T., Sandaltzopoulos, R., Becker, P., Hyman, A. & Karsenti, E. Self-organization of microtubules into bipolar spindles around artificial chromosomes in *Xenopus* egg extracts. *Nature* 382, 420–425 (1996).
- Urrutia, R., McNiven, M., Albanesi, J. P., Murphy, D. B. & Kachar, B. Purified kinesin promotes vesicle motility and induces active sliding between microtubules *in vitro*. *Proc. Natl. Acad. Sci. USA* 88, 6701–6705 (1991).
- Gyoeva, F. K., Leonova, E. V., Rodionov, V. I. & Gelfand, V. I. Intermediate filaments in fish melanophores. *J. Cell Sci.* 88, 649–655 (1987).
- Rodionov, V. I., Lim, S.-S., Gelfand, V. I. & Borisy, G. G. Microtubule dynamics in melanophores. *J. Cell Biol.* 126, 1455–1464 (1994).

Acknowledgements. We thank J. Pélouquin for preparation of rhodamine–tubulin, S. Doxsey and G. Hering for generous gifts of antibodies, and A. Verkhrvsky and P. Wilson for stimulating discussions and comments. This work was supported by a grant from the NSF.

Correspondence and requests for materials should be addressed to G.G.B. (e-mail: ggborisy@facstaff.wisc.edu).

A proton-gated cation channel involved in acid-sensing

Rainer Waldmann*, Guy Champigny*,
Frédéric Bassilana, Catherine Heurteaux
& Michel Lazdunski

Institut de Pharmacologie Moléculaire et Cellulaire, CNRS, 660 route des Lucioles, Sophia Antipolis, Valbonne 06560, France

* These authors contributed equally to this work

Acid-sensing is associated with both nociception¹ and taste transduction². Stimulation of sensory neurons by acid is of particular interest, because acidosis accompanies many painful inflammatory and ischaemic conditions. The pain caused by acids is thought to be mediated by H⁺-gated cation channels present in sensory neurons^{3–5}. We have now cloned a H⁺-gated channel (ASIC, for acid-sensing ionic channel) that belongs to the amiloride-sensitive Na⁺ channel^{6–11}/degenerin^{12–14} family of ion channels. Heterologous expression of ASIC induces an amiloride-sensitive cation (Na⁺ > Ca²⁺ > K⁺) channel which is transiently activated by rapid extracellular acidification. The biophysical and pharmacological properties of the ASIC channel closely match the H⁺-gated cation channel described in sensory neurons^{3,15,16}. ASIC is expressed in dorsal root ganglia and is also distributed widely throughout the brain. ASIC appears to be the simplest of ligand-gated channels.

The 3.5-kilobase (kb) complementary DNA we isolated from rat brain codes for a protein of 526 amino acids which shares homology with all cloned members of the amiloride-sensitive Na⁺ channel/degenerin family of ion channels (Fig. 1).

Expression of the complementary RNA in *Xenopus* oocytes induced a H⁺-gated cation channel (IH⁺) described in sensory neurons^{3,15,16}. A drop in the extracellular pH to below 6.9 activates a fast-rising and rapidly desensitizing inward current (Fig. 2a, b). The channel must be activated by extracellular protons because application of acid to the extracellular face of outside-out patches (Fig. 2c, d) activates the channel. Intracellular acidification of oocytes and acidification of the intracellular face of inside-out patches (see Methods) failed to activate the ASIC channel, nor did it alter the ASIC current induced by extracellular protons. Analysis of current–voltage (*I*–*V*) curves recorded with different extracellular cations (Fig. 2c, e) showed that Na⁺ is the major permeant ion (single-channel conductance, 14.3 pS). Like the H⁺-gated cation channel in sensory neurons^{15,16}, ASIC discriminates poorly between cations (Fig. 2c, e, f). The channel is also permeable to Li⁺, K⁺, Ca²⁺ and H⁺ with the permeability ratios pNa⁺/pLi⁺ = 1.3 (Fig. 2c, e),

$pNa^+/pK^+ = 13$ (Fig. 2c, e), $pNa^+/pCa^{2+} = 2.5$ (Fig. 2e) and $pNa^+/pH^+ = 0.8$ (Fig. 2f).

The Ca^{2+} permeability of ASIC might provide a voltage-independent pathway for Ca^{2+} entry into cells. An inward Ca^{2+} current through ASIC could be detected in the absence of extracellular Na^+ (Fig. 3a, b). The unitary conductance for Ca^{2+} was 5.2 pS (Fig. 2e). In the presence of 140 mM extracellular Na^+ , increasing external Ca^{2+} concentrations decreased the amplitude of the H^+ -activated current (Fig. 3c), demonstrating that Ca^{2+} inhibits Na^+ permeation.

A block by external Ca^{2+} is characteristic for $I(H^+)$ in sensory neurons¹⁷. The H^+ -activated inward current in sensory neurons is inhibited by amiloride¹⁸ and ethylisopropylamiloride (EIPA)¹⁹. ASIC has the same pharmacology and is reversibly blocked ($K_d = 10 \mu M$) by amiloride and its derivatives benzamil and EIPA (Fig. 3d, e).

ASIC shares 67% sequence identity with mammalian degenerin (MDEG)¹⁴ (BNaCl²⁰). However, the electrophysiological properties of those two clones expressed in *Xenopus* oocytes are different. (1)

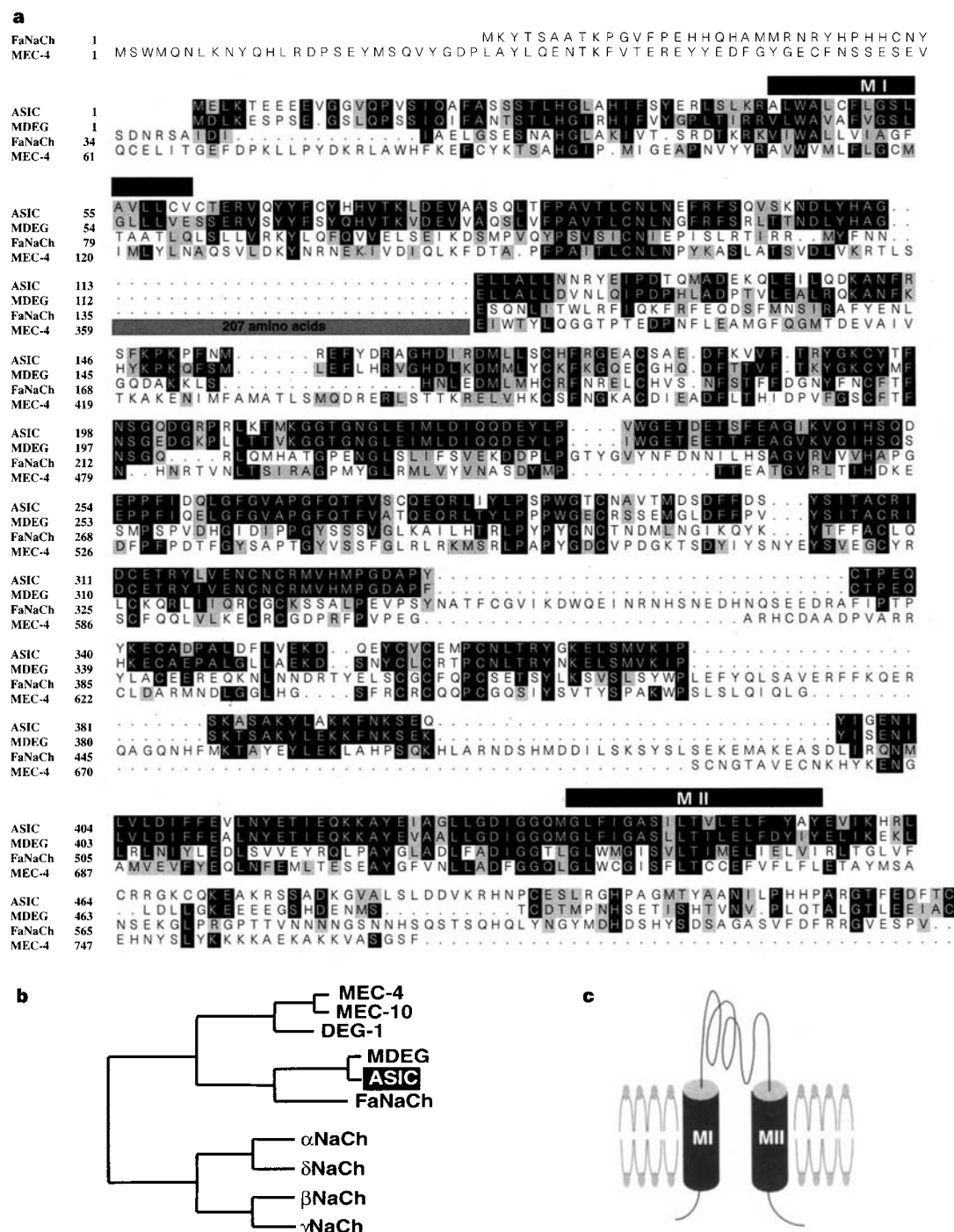


Figure 1 a, Deduced protein sequence of rat ASIC and comparison with other neuronal members of this ion-channel family: MDEG¹⁴, a mammalian neuronal cation channel activated by mutations that cause neurodegeneration with the *C. elegans* degenerins; FaNaCh¹⁰, a peptide (FMRamide)-gated Na^+ channel from *Helix aspersa*; the degenerin MEC-4¹² from *C. elegans*. Residues identical or similar to the corresponding amino acid in ASIC are printed white on black or

black on grey background, respectively. The putative transmembrane regions (MI, MII) of ASIC are indicated by black bars. **b**, Phylogenetic tree. $\alpha NaCh$, $\beta NaCh$, $\gamma NaCh$ and $\delta NaCh$ are amiloride-sensitive Na^+ -channel subunits. MEC-4, MEC-10 and DEG-1 are degenerins of *C. elegans*. **c**, Transmembrane topology proposed for this family of ion channels³⁰.

MDEG is not activated by the pH changes that activate ASIC (Fig. 2a). (2) Substitution of Gly 430 of MDEG by bulky amino acids such as Val or Phe constitutively activates the channel¹⁴, just as mutation of Ala 704 of the degenerin MEC-4 causes neurodegeneration in *Caenorhabditis elegans*¹². Identical mutations of ASIC (Gly 431 → Val; Gly 431 → Phe) do not cause any constitutive activity and the mutants cannot be activated by protons.

H⁺-gated cation channels have been described in sensory neurons but have also been recorded in neurons of the central nervous system (for review, see ref. 21). The tissue distribution of ASIC messenger RNA expression is in good agreement with this. A transcript of 4.3 kb was detected in brain by northern blot analysis (Fig. 4a) and polymerase chain reaction with reverse transcription (RT-PCR) showed that dorsal root ganglia express ASIC mRNA (Fig. 4b). ASIC mRNA is highly expressed in the small neurons of dorsal root ganglia (Fig. 5a, b), further supporting the idea that ASIC is the rapidly desensitizing H⁺-gated cation channel described in nociceptive sensory neurons. Although the presence of H⁺-gated cation channels in dorsal root ganglia is consistent with their function as acid sensors in nociception, their role in brain is unclear. *In situ* hybridization (Fig. 5c) showed a wide and heterogeneous expression of ASIC mRNA. Expression was highest in the

main olfactory bulb, cerebral cortex, hippocampal formation, habenula, basolateral amygdaloid nuclei and the cerebellum. Synaptic activity is accompanied by changes in extracellular pH^{22,23} and the rapid localized pH changes within or in the vicinity of the synaptic cleft are likely to be much steeper and higher than the macroscopic pH fluctuations reported. H⁺-gated cation channels are the only ion channels known to be directly activated by a pH change, and extracellular pH fluctuations have been proposed to have a neuromodulatory effect²³. The expression of H⁺-gated cation channels in brain further supports the hypothesis that pH fluctuations are not just a by-product of neuronal activity, but rather are a way of communicating within the central nervous system.

Besides the rapidly inactivating H⁺-gated cation channel, there are also H⁺-gated cation channels with slower kinetics in sensory neurons^{4,24}. H⁺-gated cation channels probably form, like other ligand-gated ion channels^{25,26}, a family of cation channels in which different subunits or subunit combinations form H⁺-gated cation channels with distinct pharmacological and biophysical properties.

Acid sensing is not only involved in nociception, it is also associated with taste transduction⁷. Acid stimuli activate proton-gated cation channels in taste cells^{2,27} and amiloride inhibits the perception of sour taste². Therefore both physiological and

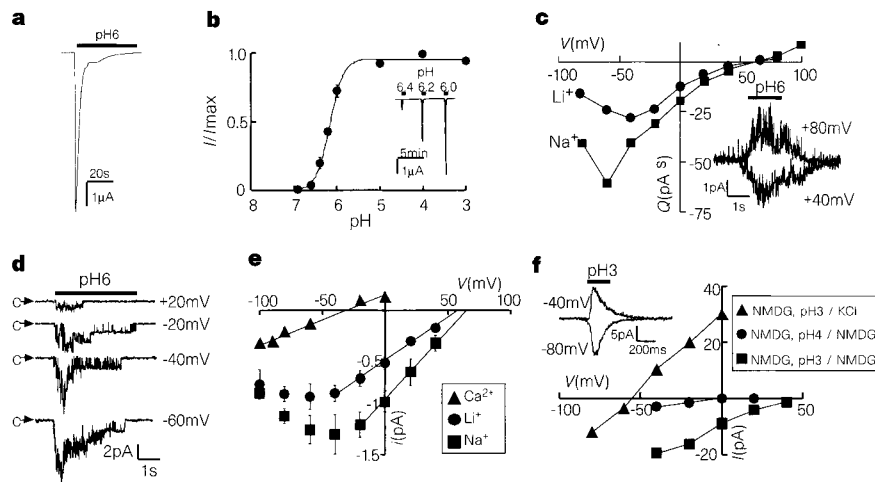


Figure 2 Properties of the H⁺-activated channel. **a**, Macroscopic inward currents recorded at -70 mV after rapid pH changes from pH 7.4 to pH 6. **b**, Dose-response curve for the extracellular pH. Initial pH was 7.4 and points represent mean values from 6 experiments; inset shows typical responses at -70 mV. **c**, Q-V relationships from outside-out patches with 140 mM Na⁺ (■) or Li⁺ (●) in the bath solution. Q is the charge transported during the acid-pH transient. Inset shows typical responses in Na⁺-containing medium. **d**, H⁺-activated currents recorded at various potentials from an outside-out patch in Na⁺-containing medium. **e**,

Mean unitary current-voltage relationships (*i*-*V*) measured from outside-out patches with 140 mM Na⁺ (■), 140 mM Li⁺ (●) or 1.8 mM Ca²⁺ (▲) as major permeant ions in the external solutions. Reversal potentials were 65, 58 and -34 mV, respectively. **f**, Proton current through the ASIC channel. Peak current-voltage relationships measured from outside-out patches in Na⁺-free, Ca²⁺-free bath solution, with pipettes containing a K⁺-free solution at pH 4 (●) and at pH 3 (■). ▲, Same conditions as (■), but with KCl in the pipette. Inset shows typical responses recorded in the condition (▲).

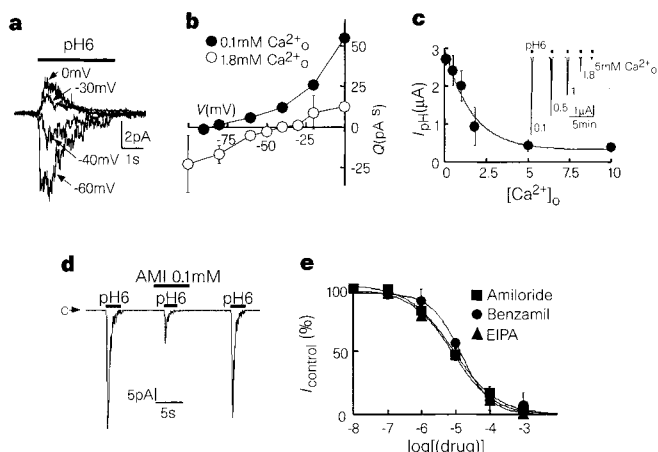


Figure 3 Effect of Ca²⁺ and amiloride on the ASIC current. **a**, H⁺-activated currents recorded at various membrane potentials from an outside-out patch with 1.8 mM Ca²⁺ in Na⁺-free bath solution. Currents reversed at -35 mV. **b**, Mean Q-V relationships from outside-out patches recorded in Na⁺-free bath solutions containing 1.8 mM Ca²⁺ (circles, reversal potential -34 mV) or 0.1 mM Ca²⁺ (filled circles, reversal potential -80 mV). **c**, Effect of external Ca²⁺ on the macroscopic peak inward current recorded at -70 mV and activated by a rapid pH change from pH 7.4 to pH 6. Inset shows typical responses. Points represent mean values ± s.e. from 5 oocytes. **d**, Effect of amiloride on H⁺-activated currents recorded at 0 mV from an outside-out patch. **e**, Inhibition of the macroscopic current (induced by a pH change from pH 7.4 to pH 6) at -70 mV by amiloride and derivatives. Points represent mean values ± s.e. from 5 oocytes.

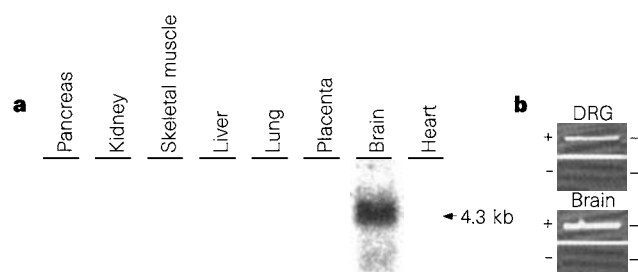


Figure 4 Tissue distribution of ASIC mRNA. **a**, Northern blot analysis of ASIC mRNA expression in human tissues. **b**, RT-PCR analysis of ASIC mRNA expression in rat brain and dorsal root ganglia (DRG). Plus and minus signs indicate samples with or without reverse transcriptase, respectively. Lanes from an ethidium bromide-stained 1% agarose gel are shown. Arrows indicate the expected size (657 bp) of the PCR product.

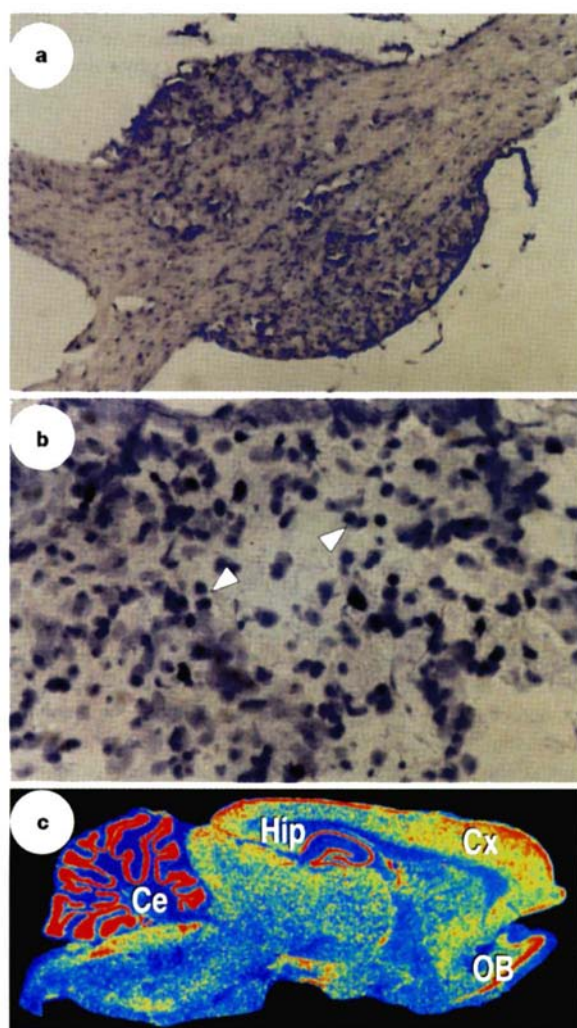


Figure 5 *In situ* hybridization. **a**, **b** Hybridization of 6- μ m sections of a dorsal root ganglion from a 3-week-old rat with the digoxigenin-labelled probe E. **a**, Low-power bright-field microphotograph (magnification, $\times 30$). **b**, High-power image (magnification, $\times 80$) of **a**. Note the intense labelling of small-diameter neurons (arrows). Similar results were also obtained with probes A, C and D. **c**, Distribution of ASIC mRNA in adult rat brain analysed by *in situ* hybridization with the antisense oligonucleotide C. Identical results were obtained with the oligonucleotide B. Colours represent transcript abundance: red, high expression; blue, not detectable). Abbreviations are: Ce, Cerebellum; Hip, Hippocampus; OB, olfactory bulb; Cx, cortex.

pharmacological data indicate that this or other members of the ASIC family of ion channels might be involved in taste transduction, although further investigation will be necessary to demonstrate this particular function.

It is particularly striking that the same class of ionic channels is associated with different facets of sensory perception: (1) amiloride-sensitive Na^+ channels are associated with the transduction of salty taste²; (2) degenerins of *C. elegans* are involved in mechanotransduction and were proposed to form mechanosensitive ion channels^{28,29}; (3) ASIC channels seem to be involved in nociception and perhaps in sour-taste transduction. The cloning of ASIC should accelerate the investigation of this group of ion channels and could help in developing specific blockers as analgesics. □

Methods

Cloning of ASIC. Conserved sequences of this ion-channel family were used to design PCR primers TTYCCIGCIRTACIITNTGYAAY (single-letter amino-acid code) and CAIARICCAIITGNCCNCCDAWRTC. A rat brain cDNA library (Stratagene) was hybridized with the 1-kb PCR product from rat brain and partial clones were isolated. The 5' end of the cDNA (202 bp) was isolated by anchored PCR after adaptor ligation to double-stranded cDNA.

Electrophysiology. *Xenopus laevis* oocytes were injected with 0.25 ng cRNA and microelectrode voltage-clamp and patch-clamp recordings were made two days after injection. Bath solutions for inside-out patch recordings and pipettes for outside-out patch and whole-cell recordings contained: 140 mM KCl (or NMDG), 2 mM MgCl_2 , 5 mM EGTA, 10 mM HEPES, pH 7.4 (with KOH). Pipettes for inside-out patch recordings and the bath solutions for outside-out and whole-cell recordings contained: 140 mM NaCl (or LiCl or NMDGCl), 2 mM MgCl_2 , 1.8 mM CaCl_2 , 10 mM HEPES, pH 7.4 (adjusted with HCl, NaOH, LiOH or tetramethylammonium hydroxide (TMAOH)). Rapid pH changes from the initial pH of 7.4 were produced by perfusion with bath solution adjusted to the pH values indicated in the figures. Intracellular acidification of oocytes was obtained by injection of 50 nl of the internal solution at pH 2 or by perfusion and removal of bath medium containing 20 mM NH_4Cl . None of the recorded currents was contaminated with the endogenous Ca^{2+} -sensitive Cl^- current of the *Xenopus* oocyte. Data were sampled at 2 kHz and filtered at 500 Hz for analysis (Biopatch software).

Northern blot, RT-PCR and *in situ* hybridization. The northern blot was purchased from Clontech and contained about 2 μ g poly(A)⁺ RNA per lane. The blot was hybridized with a random primed ³²P-labelled fragment of the partial human clone (corresponding to bases 270–764 of the rat clone) at 65 °C in 6 $\times$ SSC. For RT-PCR analysis, 5 μ g total RNA from rat brain and 3 μ g from dorsal root ganglia were reverse-transcribed and 1/30 of the sample was amplified by PCR for 30 cycles with the primers ATTGCTCTTCCCATCTCTAT and TTCAAGGCCCATACCTAAGT. Negative controls were treated in the same way except that no reverse transcriptase was added. For *in situ* hybridizations, we used antisense oligonucleotides corresponding to bases 70–114 (A), 215–248 (B), 1,821–1,859 (C), 1,896–1,940 (D) and the double-stranded DNA corresponding to bases 1,685–2,672 (E). Brain sections from adult rats were hybridized with the ³²P-end-labelled antisense oligonucleotides B or C overnight at 37 °C in 50% formamide, 2 $\times$ SSC and subsequently washed at room temperature in 1 $\times$ SSC. The signal was abolished by a 500-fold excess of unlabelled oligonucleotide. Sections of dorsal root ganglia were hybridized with oligonucleotides A, C or D labelled with digoxigenin(DIG)-dUTP and with the probe E labelled with DIG-dUTP by PCR. Probe labelling, sample preparation, hybridization and visualization of DIG nucleic acids with alkaline phosphatase-conjugated anti-DIG antibodies was carried out following the protocols from Boehringer Mannheim.

Computer analysis. The sequence alignment and the phylogenetic tree (Kimura substitution, UPGMA option) were computed with the GCG (Genetics Computer Group, Madison, Wisconsin) software package.

Received 14 November 1996; accepted 15 January 1997.

1. Rang, H. P., Bevan, S. & Dray, A. Chemical activation of nociceptive peripheral neurones. *Br. Med. Bull.* **47**, 534–548 (1991).
2. Lindemann, B. Taste reception. *Physiol. Rev.* **76**, 718–766 (1996).
3. Krishtal, O. A. & Pidoplichko, V. I. A receptor for protons in the membrane of sensory neurons may participate in nociception. *Neuroscience* **6**, 2599–2601 (1981).
4. Bevan, S. & Geppetti, P. Protons: small stimulants of capsaicin-sensitive sensory nerves. *Trends*

- Neurosci.* **17**, 509–512 (1994).
5. Akaiki, N., Krishtal, O. A. & Maruyama, T. Proton-induced sodium current in frog isolated dorsal root ganglion cells. *J. Neurophysiol.* **63**, 805–813 (1990).
 6. Canessa, C. M., Horisberger, J. D. & Rossier, B. C. Epithelial sodium channel related to proteins involved in neurodegeneration. *Nature* **361**, 467–470 (1993).
 7. Canessa, C. M. et al. Amiloride-sensitive epithelial Na⁺ channel is made of three homologous subunits. *Nature* **367**, 463–467 (1994).
 8. Lingueglia, E., Voilley, N., Waldmann, R., Lazdunski, M. & Barbry, P. Expression cloning of an epithelial amiloride-sensitive Na⁺ channel. A new channel type with homologies to *Caenorhabditis elegans* degenerins. *FEBS Lett.* **318**, 95–99 (1993).
 9. Lingueglia, E. et al. Different homologous subunits of the amiloride-sensitive Na⁺ channel are differently regulated by aldosterone. *J. Biol. Chem.* **269**, 13736–13739 (1994).
 10. Lingueglia, E., Champigny, G., Lazdunski, M. & Barbry, P. Cloning of the amiloride-sensitive FMRamide peptide-gated sodium channel. *Nature* **378**, 730–733 (1995).
 11. Waldmann, R., Champigny, G., Bassilana, F., Voilley, N. & Lazdunski, M. Molecular cloning and functional expression of a novel amiloride-sensitive Na⁺ channel. *J. Biol. Chem.* **270**, 27411–27414 (1995).
 12. Driscoll, M. & Chalfie, M. The *mec-4* gene is a member of a family of *Caenorhabditis elegans* genes that can mutate to induce neuronal degeneration. *Nature* **349**, 588–593 (1991).
 13. Huang, M. & Chalfie, M. Gene interactions affecting mechanosensory transduction in *Caenorhabditis elegans*. *Nature* **367**, 467–470 (1994).
 14. Waldmann, R., Champigny, G., Voilley, N., Lauritzen, I. & Lazdunski, M. The mammalian degenerin MDEG, an amiloride-sensitive cation channel activated by mutations causing neurodegeneration in *Caenorhabditis elegans*. *J. Biol. Chem.* **271**, 10433–10434 (1996).
 15. Kovalchuk Yu. N., Krishtal, O. A. & Nowycky, M. C. The proton-activated inward current of rat sensory neurons includes a calcium component. *Neurosci. Lett.* **115**, 237–242 (1990).
 16. Konnerth, A., Lux, H. D. & Morad, M. Proton-induced transformation of calcium channel in chick dorsal root ganglion cells. *J. Physiol.* **386**, 603–633 (1987).
 17. Davies, N. W., Lux, H. D. & Morad, M. Site and mechanism of activation of proton-induced sodium current in chick dorsal root ganglion neurones. *J. Physiol.* **400**, 159–187 (1988).
 18. Korkushko, A. & Kryshal, O. Blocking of proton-activated sodium permeability of the membranes of trigeminal ganglion neurons in the rat by organic cations. *Neurofiziol. i biokhim.* **16**, 557–561 (1984).
 19. Grantyn, R., Perouansky, M., Rodriguez-Tebar, A. & Lux, H. D. Expression of depolarizing voltage- and transmitter-activated currents in neuronal precursor cells from the rat brain is preceded by a proton-activated sodium current. *Dev. Brain Res.* **49**, 150–155 (1989).
 20. Price, M. P., Snyder, P. M. & Welsh, M. J. Cloning and expression of a novel human brain Na⁺ channel. *J. Biol. Chem.* **271**, 7879–7882 (1996).
 21. Akaiki, N. & Ueno, S. Proton-induced current in neuronal cells. *Prog. Neurobiol.* **43**, 73–83 (1994).
 22. Krishtal, O. A., Osipchuk, Y. V., Shelest, T. N. & Smirnov, S. V. Rapid extracellular pH transients related to synaptic transmission in rat hippocampal slices. *Brain Res.* **436**, 352–356 (1987).
 23. Chesler, M. & Kaila, K. Modulation of pH by neuronal activity. *Trends Neurosci.* **15**, 396–402 (1992).
 24. Bevan, S. & Yeats, J. Protons activate a cation conductance in a sub-population of rat dorsal root ganglion neurones. *J. Physiol.* **433**, 145–161 (1991).
 25. Lewis, C. et al. Coexpression of P2X₂ and P2X₃ receptor subunits can account for ATP-gated currents in sensory neurons. *Nature* **377**, 432–435 (1995).
 26. Barnard, E. A. The transmitter-gated channels: a range of receptor types and structures. *Trends Pharmacol. Sci.* **17**, 305–309 (1996).
 27. Okada, Y., Miyamoto, T. & Sato, T. Activation of a cation conductance by acetic acid in taste cells isolated from the bullfrog. *J. Exp. Biol.* **187**, 19–32 (1994).
 28. Liu, J., Schrank, B. & Waterson, R. Interaction between a putative mechanosensory membrane channel and a collagen. *Science* **273**, 361–364 (1996).
 29. Waldmann, R., Champigny, G. & Lazdunski, M. Functional degenerin-containing chimeras identify residues essential for amiloride-sensitive Na⁺ channel function. *J. Biol. Chem.* **270**, 11735–11737 (1995).
 30. Renard, S., Lingueglia, E., Voilley, N., Lazdunski, M. & Barbry, P. Biochemical analysis of the membrane topology of the amiloride-sensitive Na⁺ channel. *J. Biol. Chem.* **269**, 12981–12986 (1994).

Acknowledgements. This work was supported by the Centre National de la Recherche Scientifique (CNRS) and the Association Française contre les Myopathies (AFM). We thank J. R. de Wille and E. Lingueglia for discussion, G. Jarretou, M. Jodar and N. Leroudier for technical assistance, Y. Benhamou for secretarial assistance, and F. Aguilu for help with artwork.

Correspondence and requests for materials should be addressed to M.L. (e-mail: ipmc@unice.fr).

Ras signalling linked to the cell-cycle machinery by the retinoblastoma protein

Daniel S. Peeper[†], Todd M. Upton^{*}, Mohamed H. Ladha^{*}, Elizabeth Neuman^{*}, Juan Zalvide^{*}, René Bernards[†], James A. DeCaprio^{*} & Mark E. Ewen^{*}

^{*} The Dana-Farber Cancer Institute and the Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA

[†] The Netherlands Cancer Institute, Department of Molecular Carcinogenesis, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

The Ras proto-oncogene is a central component of mitogenic signal-transduction pathways, and is essential for cells both to leave a quiescent state (G₀) and to pass through the G₁/S transition of the cell cycle^{1–6}. The mechanism by which Ras signalling regulates cell-cycle progression is unclear, however. Here we

report that the retinoblastoma tumour-suppressor protein (Rb), a regulator of G₁ exit⁷, functionally links Ras to passage through the G₁ phase. Inactivation of Ras in cycling cells caused a decline in cyclin D1 protein levels, accumulation of the hypophosphorylated, growth-suppressive form of Rb, and G₁ arrest. When Rb was disrupted either genetically or biochemically, cells failed to arrest in G₁ following Ras inactivation. In contrast, inactivation of Ras in quiescent cells prevented growth-factor induction of both immediate-early gene transcription and exit from G₀ in an Rb-independent manner. These data suggest that Rb is an essential G₁-specific mediator that links Ras-dependent mitogenic signalling to cell-cycle regulation.

The function of Ras was neutralized to examine whether the resulting inhibition of proliferation requires Rb function. By inhibiting rather than activating Ras signalling, we could identify target proteins that are critical for, rather than proteins whose function correlates with, Ras-dependent effects on the cell cycle. Asynchronous primary *Rb*^{+/+} and *Rb*^{-/-} mouse embryo fibroblasts (MEFs) were microinjected with monoclonal antibody Y13-259, which specifically neutralizes Ras^{4,8}, and DNA replication was monitored by incorporation of 5-bromodeoxyuridine (BrdU). Neutralization of Ras led to efficient inhibition of DNA synthesis in the *Rb*^{+/+} MEFs (Fig. 1a). In contrast, inactivation of Ras led to a sevenfold smaller inhibition of DNA synthesis in *Rb*^{-/-} MEFs, a difference similar to that previously reported for microinjection of plasmids encoding p16, an inhibitor of Cdk4 and Cdk6, into both cell types^{9,10}. Ras antibody also inhibited the proliferation of immortalized 3T3 derivatives of the *Rb*^{+/+} and *Rb*^{-/-} MEFs in an Rb-dependent manner (Fig. 1b). Similar results were obtained for injections of G₁ cells after release from a serum-starved state (not shown). These results suggest that the status of Rb can dictate whether cells stop cell-cycle progression in response to inactivation of Ras.

As an independent measure to evaluate the role of Rb in Ras-dependent cell-cycle progression, we co-transfected an expression vector for a dominant interfering Ras mutant, Ras^{Asn17} (refs 11, 12), with a CD20 cell-surface marker and analysed the cell-cycle profile of the transfected population by two-colour flow cytometry. Expression of Ras^{Asn17} caused a significant arrest in the G₁ phase of the cell-cycle in NIH 3T3 cells (Fig. 2a). In contrast, expression of Ras^{Asn17} failed to cause G₁ arrest in three independent *Rb*^{-/-} 3T3 clones (Fig. 2a, b), although it was expressed to comparable levels in NIH 3T3 and *Rb*^{-/-} cells (Fig. 2a). Even higher levels of Ras^{Asn17} (obtained using stronger promoters) failed to stop *Rb*^{-/-} 3T3 cells from proliferating (data not shown). In agreement with these results, expression of Ras^{Asn17} also led to Rb-dependent inhibition of proliferation in long-term growth-suppression assays (data not shown).

To extend these observations to another cell type, we compared the effect of Ras inactivation in the *Rb*^{-/-} myoblast cell line CC42 with that in the Rb-positive myoblast cell line C2C12. Expression of Ras^{Asn17} caused significant G₁ arrest in C2C12 cells, but not in CC42 cells (Fig. 2c). Similarly, microinjection of neutralizing Ras antibodies resulted in an approximately ninefold greater inhibition of DNA synthesis in C2C12 cells than in CC42 cells (data not shown), providing further support for the results described above.

We then attempted to exclude the possibility that *Rb*^{-/-} cells might fail to respond to proliferation-restraining signals in general. Consistent with previous reports^{9,10,13,14}, p16 caused G₁ arrest in an Rb-dependent manner (Fig. 2a). In contrast, both a dominant-negative Cdk3 mutant (Cdk3-dn) and the Cdk inhibitor p27, both of which cause cell-cycle arrest in an Rb-independent manner¹⁵, induced G₁ arrest as efficiently in *Rb*^{-/-} 3T3 cells as in NIH 3T3 cells (Fig. 2a). These data indicate that the Rb requirement for induction of G₁ arrest is specific for Ras^{Asn17} and p16.

To further exclude the possibility that an Rb-independent mechanism might be responsible for the failure of *Rb*^{-/-} cells to stop proliferating in response to Ras inactivation, we performed two types of experiments. First, when Rb function was restored in *Rb*^{-/-}

cells, the ability of Ras^{Asn17} to induce G1 arrest as a function of exogenous Rb expression, and vice versa, was monitored. Ras^{Asn17} and Rb, when expressed alone, caused only a small increase in the proportion of G1 cells (Fig. 3). In contrast, in both experimental settings, coexpression of Ras^{Asn17} and Rb induced G1 arrest in a cooperative, rather than additive, fashion. Second, reversal of a Ras^{Asn17}-induced G1 arrest by the adenovirus E1A protein was largely dependent on its Rb-family binding domains (data not shown). Thus either transient inactivation or restoration of Rb can determine the response to downregulated Ras activity. It is therefore unlikely that *Rb*^{-/-} fibroblasts have acquired an adaptive mutation that renders them unresponsive to Ras signalling.

The above results suggest that, in a cycling asynchronous population of cells, inactivation of Ras causes G1 arrest in an Rb-dependent

manner. Previous work suggests that Ras activity is required for immediate-early gene activation and exit from a quiescent (G0) state^{4,8,12}. We therefore examined whether Rb is involved in either of these two processes. Ras^{Asn17} completely blocked growth-factor induction of immediate-early gene transcription in *Rb*^{-/-} 3T3 cells (Fig. 4a), as in NIH 3T3 cells¹². Moreover, inactivation of Ras prevented growth factor-stimulated cell-cycle re-entry of quiescent *Rb*^{-/-} 3T3 cells (Fig. 4b, c). These results are consistent with the observation that primary *Rb*^{-/-} MEFs and *Rb*^{-/-} 3T3 cells can be arrested by serum deprivation^{16,17} (Fig. 4). Furthermore, these data indicate that, in *Rb*^{-/-} 3T3 cells, Ras signalling can be blocked by either Ras^{Asn17} or Ras antibody. More importantly they suggest that Rb is a critical component of Ras-dependent signalling during G1 but not G0.

Our results suggest that inactivation of Ras prevents G1 exit by

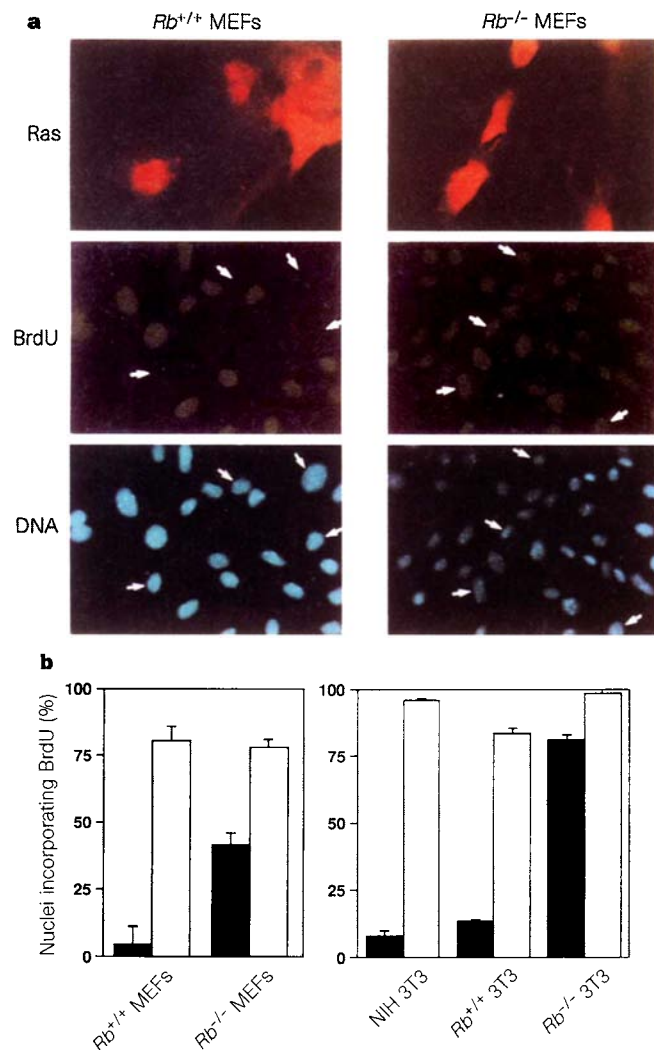


Figure 1 Microinjection of Ras-neutralizing antibody inhibits cell-cycle progression in an Rb-dependent manner. **a**, Ras-neutralizing antibody inhibits DNA synthesis in *Rb*^{+/+} MEFs but not *Rb*^{-/-} MEFs. Ras monoclonal antibody (Y13-259) was microinjected into the cytoplasm of asynchronous cultures of primary MEFs derived from *Rb*^{+/+} (left) or *Rb*^{-/-} (right) mouse embryo littermates (nuclei of injected cells are indicated by arrows). Top, staining for immunoglobulin-positive, injected cells. Middle, identical field stained for 5-bromodeoxyuridine (BrdU). Bottom, nuclear counterstaining with DAPI. **b**, Percentage of BrdU-incorporating cells after microinjection with Ras antibody (filled bars) or control IgG (open bars). As well as primary MEFs (left), immortalized (3T3) fibroblasts derived from *Rb*^{+/+} and *Rb*^{-/-} mouse embryos, and NIH 3T3 cells were used (right). At least 400 microinjected cells were assayed per cell type and antibody used. The results show mean + s.e. for at least four independent experiments.

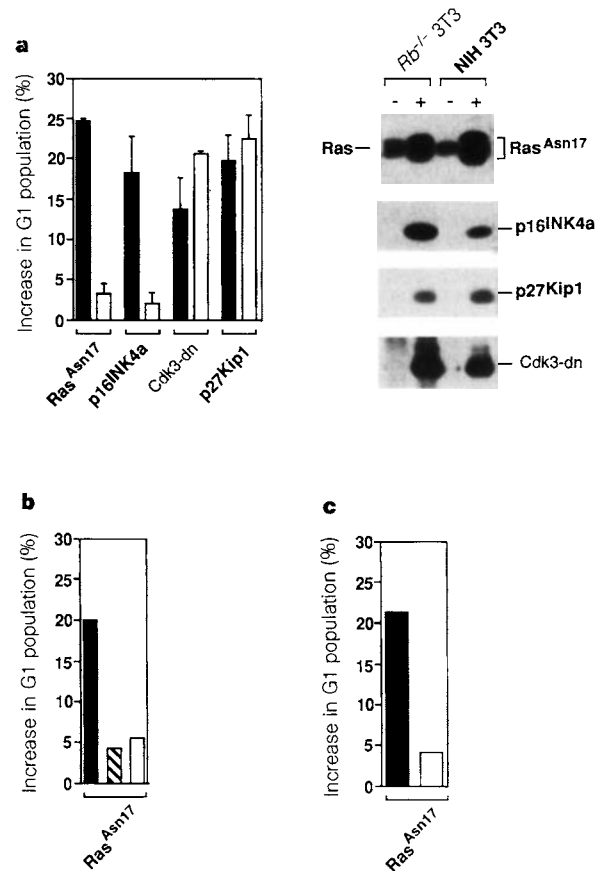


Figure 2 Expression of Ras^{Asn17} causes G1 cell-cycle arrest in an Rb-dependent manner. **a**, Left, NIH 3T3 cells (filled bars) and *Rb*^{-/-} 3T3 cells (open bars) were transfected with control plasmid, or with plasmids as indicated, together with a plasmid encoding the CD20 cell-surface marker. The transfected cell population was identified by staining with fluorescein isothiocyanate (FITC)-conjugated anti-CD20 antibody and DNA content was monitored by staining with propidium iodide and two-colour flow cytometry. The absolute changes in the percentage of cells in G1 compared to control transfections are shown with the mean + s.e. from at least three independent experiments. Right, ectopic expression of Ras^{Asn17}, p16, p27 and Cdk3-dn-HA in NIH 3T3 and *Rb*^{-/-} 3T3 cells, as visualized by immunoprecipitation, SDS-PAGE and autoradiography; — and + indicate control (vector) and expression plasmid transfections, respectively. **b**, As **a**, performed with two other independently derived immortalized *Rb*^{-/-} 3T3 cell populations (hatched and open bars), relative to NIH 3T3 cells (filled bar). A representative experiment is shown. **c**, As **a**, performed with the Rb-positive myoblast cell line C2C12 (filled bars) and the *Rb*^{-/-} myoblast cell line CC42 (open bars).

modulating Rb activity. To test this, we determined whether expression of Ras^{Asn17} affects the phosphorylation status of Rb. When expressed ectopically in NIH 3T3 cells, Rb migrated as a doublet, suggesting that both hypophosphorylated (growth-suppressive) and hyperphosphorylated (inactive) Rb species were present in roughly equimolar amounts (Fig. 5a). In contrast, when coexpressed with Ras^{Asn17}, Rb accumulated almost exclusively in its hypophosphorylated form, suggesting that inhibition of Ras activity leads to inhibition of Rb phosphorylation.

The accumulation of cyclin D proteins, the regulatory subunits of Cdk4 and Cdk6 that phosphorylate Rb, is regulated in part by mitogens¹⁸. This finding, together with the above results, prompted us to examine whether Ras inactivation decreases the accumulation of cyclin D1 protein. Indeed, a significant reduction in cyclin D1

protein levels was observed when Ras^{Asn17} was expressed (Fig. 5b, left). This effect was specific for cyclin D1, as Ras inactivation did not affect the expression of either Cdk4 or tubulin. Moreover, a Cdk2-dn-induced G1 arrest did not result in a reduction of cyclin D1. Importantly, expression of Ras^{Asn17} in *Rb*^{-/-} 3T3 cells also led to a decline in cyclin D1 (Fig. 5b, right). Because proliferating *Rb*^{-/-} 3T3 cells do not undergo G1 arrest when Ras is inactivated (Figs 1, 2 and 3), this result uncouples downregulation of cyclin D1 by Ras^{Asn17} from any cell-cycle effects, and indicates that the signalling pathway between Ras and cyclin D1 in *Rb*^{-/-} 3T3 cells is intact.

To examine whether it is the decline in cyclin D1 that causes inhibition of Rb phosphorylation when Ras is inactivated, we monitored the effect of individual G1 kinase subunits on Rb phosphorylation in this setting. Coexpression with either Cdk4 or Cdk2 alone did not have a significant effect on the inhibition of Rb phosphorylation by Ras^{Asn17} (Fig. 5a), consistent with the observation that Cdk4 levels are not altered by Ras^{Asn17} expression (Fig. 5b). In contrast, coexpression of cyclins D1, D2 or E rescued the inhibition of Rb phosphorylation by Ras^{Asn17}. This suggests that Ras inactivation causes downregulation of cyclin D1, which in turn leads to the accumulation of hypophosphorylated and active Rb, preventing G1 exit.

We then prevented the downregulation of cyclin D1 to determine whether it is an essential step during the induction of G1 arrest by Ras inactivation. Both cyclin D1-Cdk4 and cyclin D2-Cdk4, as well as activated Ras^{Val12}, completely overcame Ras^{Asn17}-induced G1 arrest, either in the absence or presence of nocodazole (Fig. 5c, d). This suggests that downregulation of cyclin D1 is required for Ras^{Asn17} to induce cell-cycle arrest. These observations are consistent with previous reports on the regulation of cyclin D1 transcription^{19,20}, that induction of cyclin D1 by oncogenic Ras may contribute to transformation²¹⁻²³, and that p16 blocks Ras plus Myc-induced transformation²⁴. Cyclin E-Cdk2 was less effective than cyclin D-Cdk4 in reversing the G1 arrest induced by Ras^{Asn17}. A possible explanation for this is the suggested coordinate action of cyclin D- and E-dependent kinases in the complete phosphorylation and inactivation of Rb²⁵. Without initial phosphorylation of Rb by cyclin D kinases, cyclin E-Cdk2 may be capable of only partly phosphorylating (and inactivating) Rb. However, this may be sufficient to cause a shift of Rb on a denaturing gel (Fig. 5a).

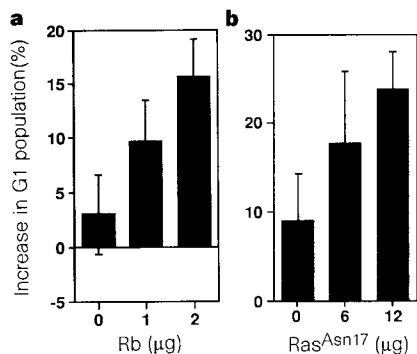


Figure 3 Reintroduction of Rb into *Rb*^{-/-} fibroblasts rescues induction of G1 arrest by Ras inactivation. **a**, Individual cell populations of *Rb*^{-/-} 3T3 cells were transfected with a constant amount of plasmids expressing Ras^{Asn17} and CD20, together with an increasing amount of a plasmid encoding Rb. After transfection, cells were processed and the absolute change in the percentage of cells in G1, relative to a control transfection, was determined by FACS analysis. Data were corrected for cell-cycle effects caused by expression of Rb alone; mean + s.e. for four independent experiments. **b**, As **a**, except Rb plasmid was held constant and Ras^{Asn17} expression plasmid was added in increasing amounts. Cell-cycle effects caused by expression of Ras^{Asn17} in the absence of Rb, were corrected for; mean + s.e. for two independent experiments.

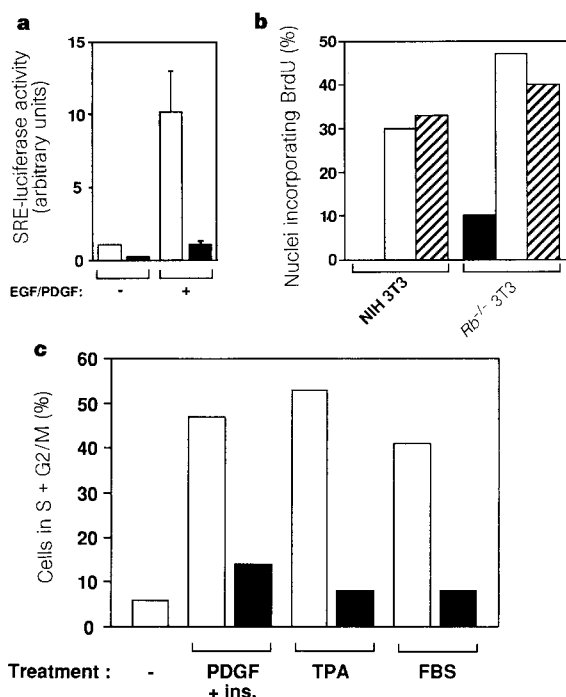


Figure 4 Inhibition of Ras activity blocks G0 exit in an Rb-independent manner. **a**, Ras^{Asn17} expression inhibits growth factor-induced serum response element (SRE) activity in *Rb*^{-/-} 3T3 cells. *Rb*^{-/-} 3T3 cells were transfected with Ras^{Asn17} expression plasmid (filled bars) or empty vector (open bars), together with an SRE-luciferase reporter plasmid. Then, 24 h after transfection, cells were serum starved (0.1% FBS) for 72 h. At this time, no growth factors or PDGF and EGF (10 ng ml⁻¹ each) were added to the cultures and luciferase activity was measured 24 h later; mean + s.e. for three independent experiments are shown. Similar results were obtained for NIH 3T3 cells (not shown). **b**, Ras neutralization inhibits growth factor-induced G0 exit in *Rb*^{-/-} 3T3 cells. Cells rendered quiescent by serum starvation were microinjected with either control (rabbit IgG; open bars) or anti-Ras antibody (filled bars), stimulated with EGF and PDGF (10 ng ml⁻¹ each) and analysed for incorporation of BrdU. The absolute percentages of cells incorporating BrdU for each cell type under each condition are shown. Percentages are for over 150 injected cells from at least two independent experiments. The absolute percentage of uninjected cells incorporating BrdU after growth factor stimulation is indicated (hatched bars). **c**, Ras^{Asn17} expression inhibits growth factor-induced G0 exit in *Rb*^{-/-} 3T3 cells. *Rb*^{-/-} 3T3 cells were serum starved for 24 h, infected with recombinant Ras^{Asn17} adenovirus (open bars) or control adenovirus (filled bars) as indicated, and starved for another 16 h. Cells were treated with either PDGF (10 ng ml⁻¹) + insulin (ins., 10 μg ml⁻¹), TPA (25 nM) or FBS (2%) and cell-cycle profiles were determined by FACS analysis 22 h later. The absolute percentages of cells in S + G2/M phases are shown. The control (left lane) shows the percentage of cells in S + G2/M just before stimulation.

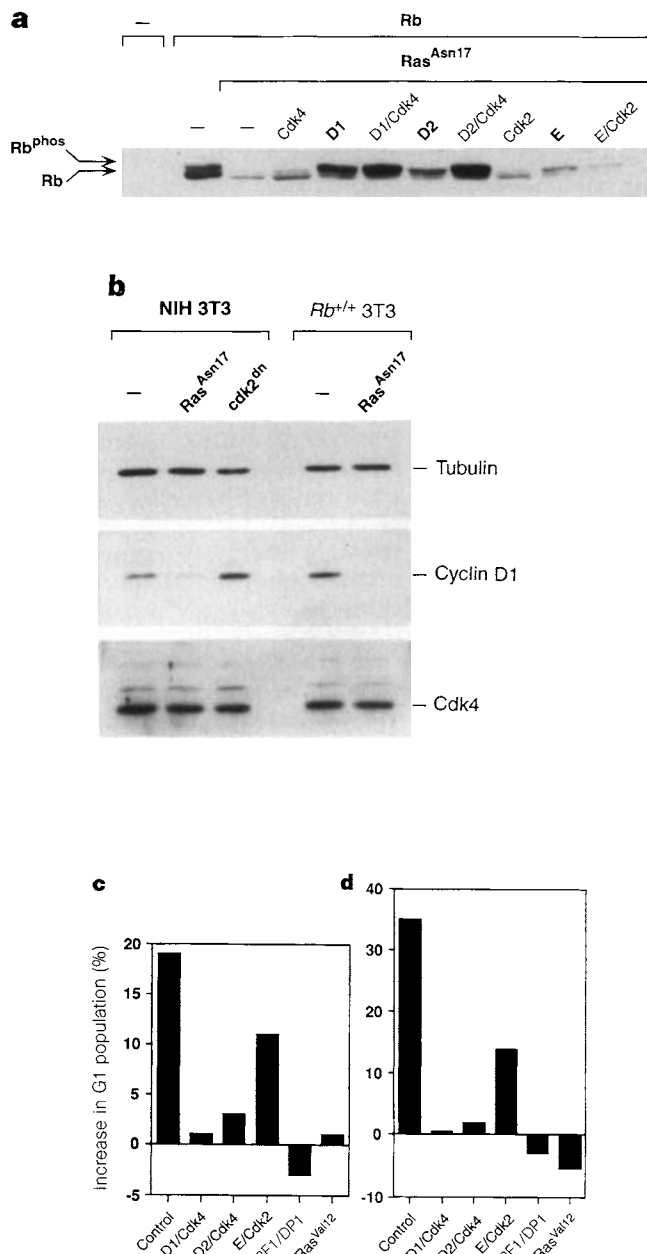


Figure 5 D-type cyclins and E2F overcome the inhibitory effect of Ras inactivation on cell-cycle progression. **a**, NIH 3T3 cells were transfected with plasmid encoding human Rb, either in the absence or presence of expression vectors for Ras^{Asn17} and cyclin-dependent kinase subunits, as indicated. Rb was detected by western blotting. **b**, NIH 3T3 and Rb^{+/+} 3T3 cells were transfected with plasmids encoding either Ras^{Asn17}, a dominant-negative Cdk2 mutant or empty vector, as indicated, together with a plasmid encoding the CD20 cell-surface marker. Extracts from transfected cells isolated by CD20-specific cell sorting were subjected to western blotting with specific antibodies, as indicated. **c**, NIH 3T3 cells were co-transfected with plasmids encoding CD20, Ras^{Asn17} and plasmids encoding cyclin-Cdk kinases, E2F1 and DP1, or Ras^{Val12}, as indicated. As a control, CD20 and Ras^{Asn17} plasmids were co-transfected with empty vector (first lane). The cell-cycle profiles of the transfected cell populations were determined by FACS analysis. A representative experiment from at least three independent experiments, with the absolute change in percentage of cells in G1 compared to control transfections, is shown. **d**, As **c**, except that nocodazole (50 ng ml⁻¹) was added to the cultures 24 h after transfection to reduce the background proportion of cells in G1.

Our results predict that downstream components within the Rb pathway should also interfere with the inhibition of proliferation when Ras is inactivated. The E2F1-DP1 transcription factor, an Rb-effector, completely overcame Ras^{Asn17}-induced G1 arrest (Fig. 5c, d), thus providing additional support for a model in which cyclin D, Rb and E2F are components of a critical Ras-effector pathway. However, more complex signalling networks may be involved in Ras-mediated transformation. Nevertheless, our results suggest that Rb, by acting in a pathway with Ras, is important for adequate growth regulation during antimitogenic Ras-dependent signalling. □

Methods

Microinjection. MEFs isolated from Rb^{+/+} and Rb^{-/-} mouse embryo littermates and their immortalized 3T3 derivatives²⁶ were cultured in DMEM with 10% FBS. NIH 3T3 cells were cultured in DMEM containing 5% bovine calf serum. Cells were plated onto photo-etched coverslips preincubated in PBS–0.1% gelatin. Asynchronous cells were microinjected with either Ras monoclonal antibody Y13-259 (Oncogene Science) or rabbit IgG (Sigma) at 5 mg ml⁻¹ in PBS. At least 100 cells were injected in each experiment. BrdU (10 μM) was added 11–13 h post-injection and processed 24–26 h post-injection. To process microinjected cells for immunofluorescence, coverslips were rinsed twice with PBS and fixed for 10 min with 3% paraformaldehyde in PBS–2% sucrose. Coverslips were then rinsed three times in PBS and cells were permeabilized in PBS–0.2% Triton X-100. Coverslips were then rinsed three times in PBS and blocked for 15–20 min with PBS containing 0.1% Triton X-100 and 3% BSA. This and subsequent washes were done three times in PBS–0.1% Triton X-100. Coverslips were then incubated with either rhodamine-conjugated goat anti-rabbit or rhodamine-conjugated goat anti-rat secondary antibody, diluted 1:200 in blocking buffer, in a humidified chamber for 15 min. After washing, slips were incubated 10 min in 1.5 M HCl. Coverslips were washed and incubated for 30 min in a humidified chamber with anti-BrdU antibody, diluted 1:2.5 in PBS–0.1% Triton X-100. Coverslips were washed and nuclei were stained with DAPI. For analysis of quiescent fibroblasts, cells were maintained in 0.1% FBS for 48 h before microinjection with either control rabbit IgG or anti-Ras antibody. Cells were stimulated with EGF and PDGF 2–4 h post-injection and BrdU was added 8–10 h later. Cells were processed for immunofluorescence 24 h post-injection.

Cell-cycle analysis. To determine cell-cycle effects by Ras^{Asn17} expression, 4 μg pCMV-CD20 was co-transfected with 21 μg of either pMT-Ras^{Asn17} (gift from G. Cooper) or control plasmid pMT-ΔBam into cycling Rb^{-/-} 3T3, NIH 3T3, CC42 (ref. 27; gift from J. Schneider) and C2C12 cells. As controls, 21 μg of either pcDNA3-p16, pCMV-Cdk3-dn-HA (gift from E. Harlow), or pCMV5-p27 (gift from J. Massagué) was co-transfected into Rb^{-/-} 3T3 and NIH 3T3 cells. For fluorescence-activated cell sorting (FACS) analysis, CD20-co-transfected cells were identified with CD20-specific FITC-conjugated antibodies 44 h post-transfection²⁸. After propidium iodide staining of DNA, the cell-cycle profiles (2,000–10,000 CD20⁺ cells per sample) were analysed by FACS on a Becton-Dickinson FACScan machine. For cell-cycle analysis of quiescent Rb^{-/-} 3T3 cells upon growth-factor stimulation, cells were starved in 0.1% FBS for 24 h, infected with recombinant adenoviruses, starved for another 16 h, stimulated for 22 h, and processed. To generate a control plasmid for Ras^{Asn17}, pMT-ΔBam, the entire Ras^{Asn17} coding sequence was removed from pMT-Ras^{Asn17} and the vector was religated. pcDNA3-p16 was generated by inserting an EcoRI–Xho fragment from pBS-p16 (ref. 29; gift from D. Beach) into pcDNA3 (Invitrogen). In the Rb restoration experiments, Rb^{-/-} 3T3 cells were transfected with 4 μg pCMV-CD20, 12 μg of either pMT-Ras^{Asn17} or pMT-ΔBam, and pCMV-Rb (0, 1 or 2 μg). Alternatively, pCMV-Rb (4 μg) was cotransfected with increasing amounts of pMT-Ras^{Asn17} (0, 6 or 12 μg). To rescue Ras^{Asn17}-induced G1 arrest, 2 μg pCMV-CD20 was co-transfected with 9 μg of either pMT-Ras^{Asn17} or pMT-ΔBam. The following plasmids were co-transfected: 10 μg of cyclin plasmid with 4 μg of Cdk plasmid, 10 μg pCMV-E2F1 with 4 μg pCMV-DP1, 14 μg of pPCR-Hras^{Val12}, or pcDNA3. To generate recombinant Ras^{Asn17} adenovirus, Ras^{Asn17} coding sequence was cloned into pCA13 and co-transfected with pBH10, containing part of the adenoviral genome, into helper cells (Microbix Biosystems, Toronto, Canada).

Immunoprecipitation and western blotting. Immunoprecipitation²⁸ was

performed with either monoclonal Ras antibody Y13-259, HA monoclonal antibody 12CA5 for Cdk3-dn, p16 monoclonal antibody ZJ11 (Upstate Biotechnology) or p27 polyclonal antibody (M-197, Santa Cruz). To determine Rb phosphorylation status, 8 µg of pCMV-Rb was co-transfected with 8 µg of either pMT-Ras^{Asn17} or pMT-ΔBam. Plasmids (6 µg) for cyclins D1, D2 (ref. 28) or E (gift from R. Weinberg) subcloned into pRc/CMV (Invitrogen), or 3 µg for Cdk (pCMV-Cdk4-HA or pCMV-Cdk2 (gift from E. Harlow)) plasmid were co-transfected. To determine the effect of Ras^{Asn17} on specific cellular proteins, 5 µg of pCMV-CD20 was co-transfected with 5 µg pCDNA3 and 15 µg of pMT-Ras^{Asn17}, pMT-ΔBam or pCMV-Cdk2-dn (gift from E. Harlow). For cell sorting, cells were processed 44 h after transfection, and the CD20-positive population was isolated on a Becton-Dickinson Cell Sort machine, lysed and used for western blotting (80,000 cells) with antibodies; Rb (C15, Santa Cruz), cyclin D1 (DSC-6, Sanbyo), rabbit antiserum for Cdk4 (gift from C. Sherr), or anti-tubulin antibody YLI/2 (ECACC).

Luciferase assays. Cells were co-transfected with 5 µg pSRE-luc (gift from R. Davis) and either 15 µg pMT-Ras^{Asn17} or pMT-ΔBam, and 5 µg CMV-βgal. Cells were incubated 20 h post-transfection in DMEM containing 0.1% FBS. After 3 days, cells were stimulated with EGF and PDGF (R&D Systems), and luciferase activity was measured 24 h later³⁰.

Received 29 August 1996; accepted 10 January 1997.

- Pardee, A. B. G1 events and regulation of cell proliferation. *Science* **246**, 603–608 (1989).
- Pronk, G. J. & Bos, J. L. The role of p21ras in receptor tyrosine kinase signalling. *Biochim. Biophys. Acta* **1198**, 131–147 (1994).
- Marshall, C. J. Specificity of receptor tyrosine kinase signalling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80**, 179–185 (1995).
- Dobrowolski, S., Harter, M. & Stacey, D. W. Cellular ras activity is required for passage through multiple points of the G0/G1 phase in Balb/c 3T3 cells. *Mol. Cell. Biol.* **14**, 5441–5449 (1994).
- Feramisco, J. R., Gross, M., Kamata, T., Rosenberg, M. & Sweet, R. W. Microinjection of the oncogene form of the human H-ras (T-24) protein results in rapid proliferation of quiescent cells. *Cell* **38**, 109–117 (1984).
- Stacey, D. W. & Kung, H. F. Transformation of NIH 3T3 cells by microinjection of Ha-ras protein. *Nature* **310**, 508–511 (1984).
- Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
- Mulcahy, L. S., Smith, M. R. & Stacey, D. W. Requirement for ras proto-oncogene function during serum-stimulated growth of NIH 3T3 cells. *Nature* **313**, 241–243 (1985).
- Lukas, J. *et al.* Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. *Nature* **375**, 503–506 (1995).
- Koh, J., Enders, G. H., Dynlacht, B. D. & Harlow, E. Tumour-derived p16 alleles encoding proteins defective in cell-cycle inhibition. *Nature* **375**, 506–510 (1995).
- Feig, L. A. & Cooper, G. M. Inhibition of NIH 3T3 cell proliferation by a mutant ras protein with preferential affinity for GDP. *Mol. Cell. Biol.* **8**, 3235–3243 (1988).
- Cai, H., Szeberenyi, J. & Cooper, G. M. Effect of a dominant inhibitory Ha-ras mutation on mitogenic signal transduction. *Mol. Cell. Biol.* **10**, 5314–5323 (1990).
- Medema, R. H., Herrera, R. E., Lam, F. & Weinberg, R. A. Growth suppression by p16ink4 requires functional retinoblastoma protein. *Proc. Natl. Acad. Sci. USA* **92**, 6289–6293 (1995).
- Guan, K.-L. *et al.* Growth suppression by p18, a p16INK4/MTS1- and p14INK4B/MTS2-related CDK6 inhibitor, correlates with wild-type pRb function. *Genes Dev.* **8**, 2939–2952 (1994).
- Sherr, C. J. & Roberts, J. M. Inhibitors of mammalian G1 cyclin-dependent kinases. *Genes Dev.* **9**, 1149–1163 (1995).
- Herrera, R. E. *et al.* Altered cell cycle kinetics, gene expression, and G1 restriction point regulation in Rb-deficient fibroblasts. *Mol. Cell. Biol.* **16**, 2402–2407 (1996).
- Lukas, J., Bartkova, J., Rohde, M., Strauss, M. & Bartek, J. Cyclin D1 is dispensable for G1 control in retinoblastoma gene-deficient cells independently of cdk4 activity. *Mol. Cell. Biol.* **15**, 2600–2611 (1995).
- Sherr, C. J. Mammalian G1 cyclins. *Cell* **73**, 1059–1065 (1993).
- Lavoie, J. N., L'Allemand, G., Brunet, A., Muller, R. & Pouyssegur, J. Cyclin D1 expression is regulated positively by the p42/p44MAPK and negatively by the p38/HOGMAPK pathway. *J. Biol. Chem.* **271**, 20608–20616 (1996).
- Albanese, C. *et al.* Transforming p21ras mutants and c-Ets-2 activate the cyclin D1 promoter through distinguishable regions. *J. Biol. Chem.* **270**, 23589–23597 (1995).
- Films, J. *et al.* Induction of cyclin D1 overexpression by activated ras. *Oncogene* **9**, 3627–3633 (1994).
- Liu, J.-J. *et al.* Ras transformation results in an elevated level of cyclin D1 and acceleration of G1 progression in NIH 3T3 cells. *Mol. Cell. Biol.* **15**, 3654–3663 (1995).
- Winston, J. T., Coats, S. R., Wang, Y.-Z. & Pledger, W. J. Regulation of the cell cycle machinery by oncogenic Ras. *Oncogene* **12**, 127–134 (1996).
- Serrano, M., Gomez-Lohoz, E., DePinho, R. A., Beach, D. & Bar-Sagi, D. Inhibition of Ras-induced proliferation and transformation by p16INK4. *Science* **267**, 249–252 (1995).
- Hatakeyama, M., Brill, J. A., Fink, G. R. & Weinberg, R. A. Collaboration of G1 cyclins in the functional inactivation of the retinoblastoma protein. *Genes Dev.* **8**, 1759–1771 (1994).
- Zalvide, J. & DeCaprio, J. A. Role of pRb-related proteins in simian virus 40 large-T-antigen-mediated transformation. *Mol. Cell. Biol.* **15**, 5800–5810 (1995).
- Schneider, J. W., Gu, W., Zhu, L., Mahdavi, V. & Nadal-Ginard, B. Reversal of terminal differentiation mediated by p107 in Rb^{-/-} muscle cells. *Science* **264**, 1467–1471 (1994).
- Ewen, M. E. *et al.* Functional interactions of the retinoblastoma protein with mammalian D-type cyclins. *Cell* **73**, 487–497 (1993).
- Serrano, M., Hannon, G. J. & Beach, D. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/cdk4. *Nature* **366**, 704–707 (1993).
- Ewen, M. E., Oliver, C. J., Sluss, H. K., Miller, S. J. & Peeper, D. S. p53-dependent repression of CDK4 translation in TGF-β-induced cell-cycle arrest. *Genes Dev.* **9**, 204–217 (1995).

Acknowledgements. D.S.P. and T.M.U. contributed equally to this work. We thank R. Medema for critical ideas; D. Livingston, P. Adams, D. Ginsberg, R. Scully, S. Bhattacharya, W. Sellers and our colleagues in the Ewen and Bernards laboratories for criticism and encouragement; P. Adams, D. Beach, G. Cooper,

R. Davis, T. Ernst, E. Harlow, P. Hinds, P. Hordijk, T. Jacks, W. Kaelin, O. Kranenburg, W. Krek, J. Massagué, H. Masselink, J. Schneider, C. Sherr, M. Voorhoeve, R. Weinberg, E. White, Y. Xiong and R. Zwijsen for plasmids, antibodies, cell lines, growth factors, mice and help with generating recombinant adenovirus; and members of the flow-cytometry facilities for their help. D.S.P. thanks Y. C. Peeper for support. D.S.P. is a fellow of the Dutch Cancer Society. E.N. is supported by a fellowship from the Woman's Cancer Program at the Dana-Farber Cancer Institute. This work was supported by grants from the NIH and the Sandoz/Dana-Farber Drug Discovery Program (M.E.E.).

Correspondence and requests for materials should be addressed to M.E.E. (e-mail: mark_ewen@dfci.harvard.edu).

A family of proteins that inhibit signalling through tyrosine kinase receptors

Alexei Kharitonov^{*†}, Zhengjun Chen^{*†}, Irmi Sures[‡], Hongyang Wang[‡], James Schilling[‡] & Axel Ullrich[†]

[†] Department of Molecular Biology, Max-Planck-Institute für Biochemie, Am Klopferspitz 18A, 82152 Martinsried, Germany

[‡] Sugen Inc., 515 Galveston Road, Redwood City, California 94063, USA

^{*} These authors contributed equally to this work

Phosphotyrosine phosphatases are critical negative or positive regulators in the intracellular signalling pathways that result in growth-factor-specific cell responses such as mitosis, differentiation, migration, survival, transformation or death^{1–4}. The SH2-domain-containing phosphotyrosine phosphatase SHP-2 is a positive signal transducer for several receptor tyrosine kinases (RTKs) and cytokine receptors^{5–7}. To investigate its mechanism of action we purified a tyrosine-phosphorylated glycoprotein which in different cell types associates tightly with SHP-2 and appears to serve as its substrate. Peptide sequencing in conjunction with complementary DNA cloning revealed a new gene family of at least fifteen members designated signal-regulatory proteins (SIRPs). They consist of two subtypes distinguished by the presence or absence of a cytoplasmic SHP-2-binding domain. The transmembrane polypeptide SIRPα1 is a substrate of activated RTKs and in its tyrosine-phosphorylated form binds SHP-2 through SH2 interactions and acts as its substrate. It also binds SHP-1 and Grb2 *in vitro* and has negative regulatory effects on cellular responses induced by growth factors, oncogenes or insulin. Our findings indicate that proteins belonging to the SIRP family generally regulate signals defining different physiological and pathological processes.

Analysis of anti-SHP-2 immunoprecipitates from both pervanadate (POV) as well as growth-factor-stimulated cells revealed a major tyrosine-phosphorylated protein. In mouse mammary tumour (MM5/C1) cells, Rat1 cells overexpressing the human insulin receptor (Rat1-IR) and human epidermoid carcinoma (A431) cells, this phosphoprotein migrated at positions corresponding to relative molecular masses of 120K, 110K and 90K, respectively (Fig. 1a), which upon *in vitro* deglycosylation was in all cases reduced to 65K. This indicated that the same 65K SHP-2-binding protein was differentially glycosylated in a species-specific manner. In some cell lines, for example in A431 cells, other tyrosine-phosphorylated proteins in the 90K–120K range were unaffected by the deglycosylation treatment (Fig. 1b) and may in fact be Gab1 (ref. 8) and/or the human homologue of the *Drosophila* DOS protein⁹. Insulin-treated Rat1-IR cells were used to purify the 110K SHP-2-binding glycoprotein (Fig. 1b, right) by standard chromatography procedures (see Methods). About 4 µg of the semi-pure glycoprotein that copurified with SHP-2 (Fig. 1c) was subjected to microsequence analysis: this yielded five peptide sequences. Computer-aided search for the corresponding DNAs in the expressed-sequence tag (EST) database led to the identification of a 305-base-pair (bp) rat sequence (accession number H31804) and of a human cDNA fragment of 2 kilobases (kb) (EMBL database,

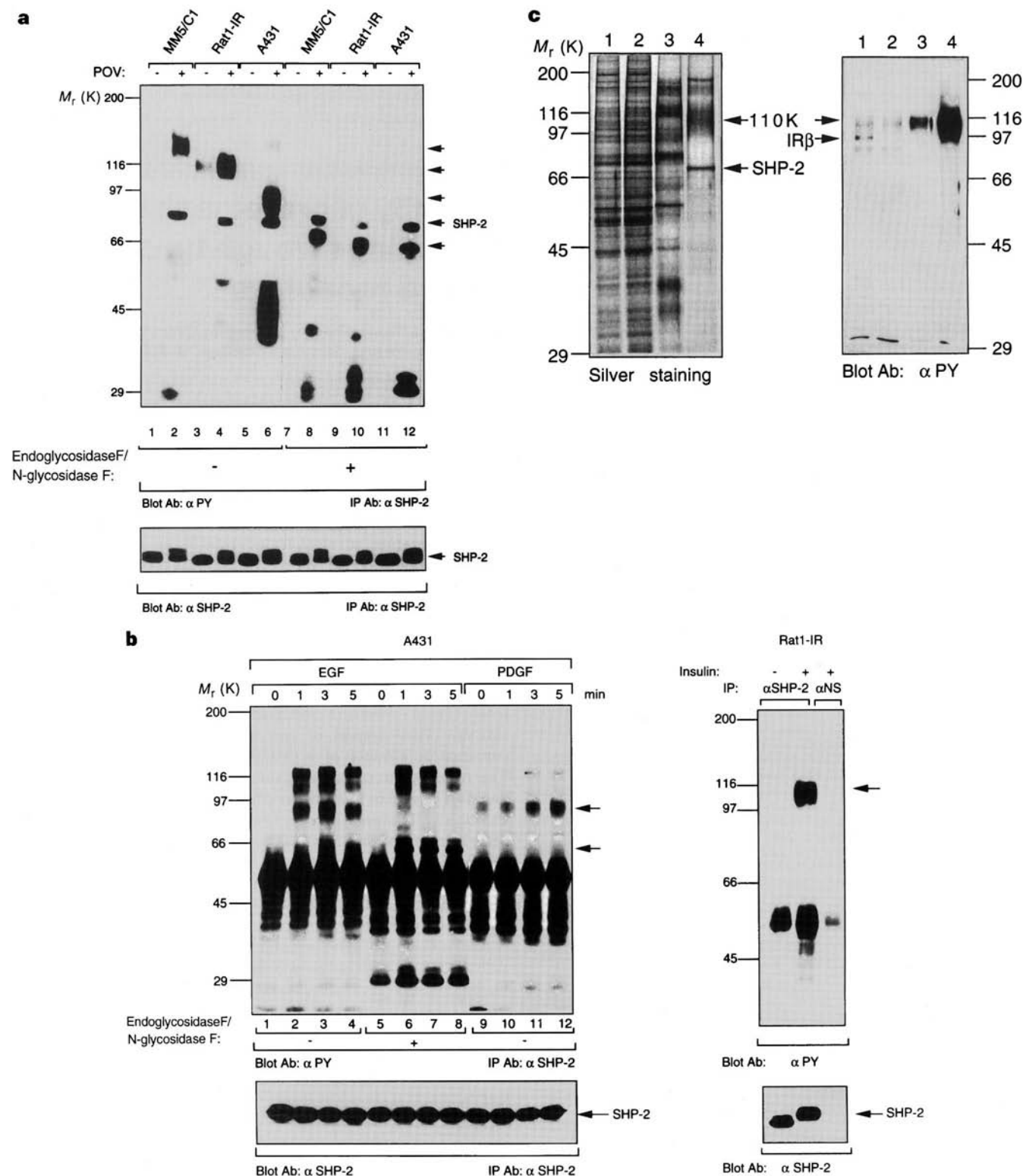


Figure 1 Identification and purification of tyrosine-phosphorylated SHP-2-associated proteins. **a**, Western blots containing anti-SHP-2 immunoprecipitates from starved or POV-treated mouse MM5/C1 mammary carcinoma (lanes 1, 2, 7 and 8), rat fibroblast Rat1-IR (lanes 3, 4, 9 and 10) or human epidermal carcinoma A431 (lanes 5, 6, 11 and 12) cells were incubated with monoclonal anti-phosphotyrosine antibodies (top) and monoclonal anti-SHP-2 (bottom) antibodies. Arrows indicate the relative positions of the 90K–120K tyrosine phosphorylated proteins that had not been treated with endoglycosidase F/N-glycosidase F (EndoF/F) (lanes 1–6) or after *in vitro* deglycosylation (lanes 7–12). **b**, Western blots containing anti-SHP-2 immunoprecipitates from quiescent, EGF- or PDGF-stimulated A431 cells, and starved or insulin-treated Rat1-IR cells were incubated

with anti-phosphotyrosine antibodies (top) or anti-SHP-2 (bottom) antibodies. Samples were deglycosylated with (lanes 5–8) or treated without EndoF/F (lanes 1–4 and 9–12). As a control, insulin-stimulated Rat1-IR cell lysates were immunoprecipitated with preimmune rabbit serum (α NS). Plain arrows indicate the relative position of the tyrosine-phosphorylated proteins with or without *in vitro* deglycosylation. **c**, SDS-PAGE analysis of the purified 110K protein. Samples from each purification step (lanes: 1, solubilized crude membrane extract; 2, hIR-depleted extract; 3, concentrated eluate from WGA-agarose beads; 4, eluate from anti-phosphotyrosine antibody column) were analysed by 10% SDS-PAGE and visualized by silver staining (left) and in western blots using monoclonal anti-phosphotyrosine antibodies (right).

accession number U6701) containing matching and homologous sequences, respectively. Specific primers flanking the 5'-most portion of this sequence were used to amplify a 360-bp human DNA fragment (encoding Ex1 in Fig. 2c) which was used to screen several human cDNA libraries.

Characterization of a number of independent clones yielded four complete and distinct sequences (Fig. 2a, b) encoding proteins of two structural subtypes designated SIRP α and β . The deduced 503-amino-acid (M_r 57,000 (57K)) SIRP α 1 sequence contains a putative 24-residue transmembrane domain which separates an extracellular portion with three immunoglobulin-like structures from a cyto-

plasmic region that includes four potential tyrosine-phosphorylation sites and one proline-rich region. SIRP β 1, which although it is highly homologous to the α -form within the immunoglobulin-like domains (Fig. 2a), lacks the cytoplasmic portion but still contains at the carboxy terminus 24 residues that are highly hydrophobic and might function as a membrane-spanning anchor. Preliminary chromosomal analysis indicates that the β -form of the SIRP family is not a result of alternative splicing (results not shown).

Figure 2b shows the complete sequences of two additional members of the SIRP α family: α 2 exhibits sequence divergence only in the first immunoglobulin-like domain, whereas α 3 has amino-acid substitutions throughout the entire polypeptide.

In addition to the four SIRP α and SIRP β sequences shown in Fig. 2a and b, eleven closely related sequences of the first immunoglobulin-like domain shown in Fig. 2c provide evidence for the existence of a large new family of proteins. Moreover, the diversity of SIRP family members is extended by alternative splicing of individual genes (results not shown). Further evidence for this diversity is provided by northern blot analysis using a SIRP α -type-specific probe (Fig. 2d): in addition to several weaker bands, two major RNA transcripts of 2.5 and 3.9 kb were detected in all tissues examined, with expression being highest in brain, liver and heart. But, given the extent of sequence homology, the question of tissue-specific SIRP α 1 subtype expression still needs investigation.

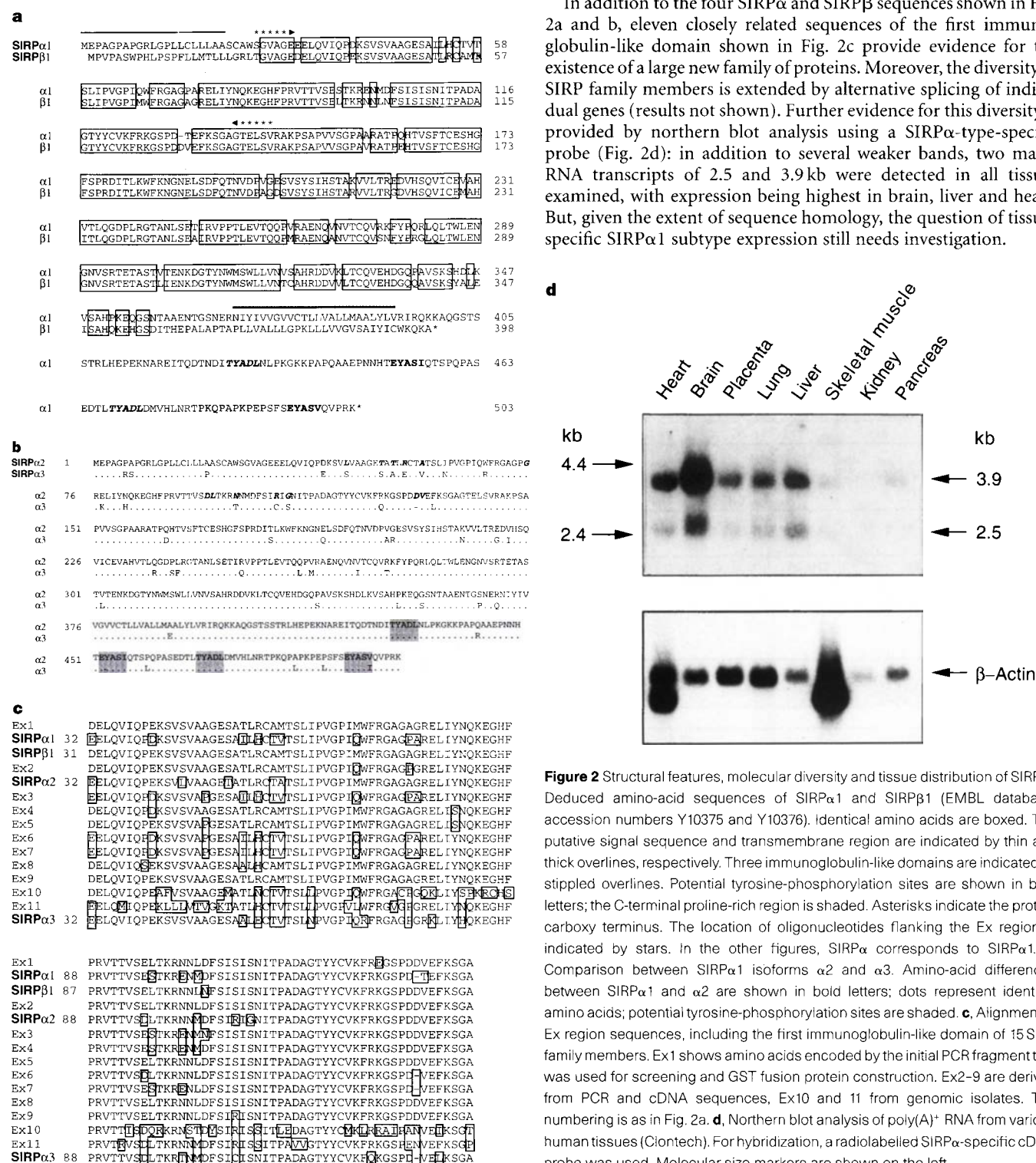


Figure 2 Structural features, molecular diversity and tissue distribution of SIRP. **a**, Deduced amino-acid sequences of SIRP α 1 and SIRP β 1 (EMBL database, accession numbers Y10375 and Y10376). Identical amino acids are boxed. The putative signal sequence and transmembrane region are indicated by thin and thick overlines, respectively. Three immunoglobulin-like domains are indicated by stippled overlines. Potential tyrosine-phosphorylation sites are shown in bold letters; the C-terminal proline-rich region is shaded. Asterisks indicate the protein carboxy terminus. The location of oligonucleotides flanking the Ex region is indicated by stars. In the other figures, SIRP α corresponds to SIRP α 1. **b**, Comparison between SIRP α 1 isoforms α 2 and α 3. Amino-acid differences between SIRP α 1 and α 2 are shown in bold letters; dots represent identical amino acids; potential tyrosine-phosphorylation sites are shaded. **c**, Alignment of Ex region sequences, including the first immunoglobulin-like domain of 15 SIRP family members. Ex1 shows amino acids encoded by the initial PCR fragment that was used for screening and GST fusion protein construction. Ex2-9 are derived from PCR and cDNA sequences, Ex10 and 11 from genomic isolates. The numbering is as in Fig. 2a. **d**, Northern blot analysis of poly(A)⁺ RNA from various human tissues (Clontech). For hybridization, a radiolabelled SIRP α -specific cDNA probe was used. Molecular size markers are shown on the left.

The identification of SIRP α 1 as an SHP-2-binding protein and substrate was confirmed by expressing the cDNA either alone or in combination with vectors that express SHP-2 or an inactive mutant (SHP-2C463A¹⁰) in BHK cells which stably express receptors for human epidermal growth factor (EGF), insulin or β -platelet-derived growth factor (PDGF). For immunoprecipitation, a polyclonal antibody raised against a glutathione-S-transferase (GST) fusion protein containing the extracellular Ex1 region was used (Fig. 2b). Anti-Ex1 immunoprecipitates revealed a tyrosine-phosphorylated protein of 85K–90K which associated with SHP-2 upon ligand stimulation. Our results indicated that SIRP α 1 was a substrate for SHP-2 because expression of SHP-2C463A led—even in starved cells—to a significant increase in its phosphotyrosine content, whereas coexpression with wild-type SHP-2 resulted in its dephosphorylation (Fig. 3a). Tyrosine-phosphorylated SIRP α 1 was also detected in anti-SHP-2 immunoprecipitates, confirming the stable complex formation between SIRP α 1 and SHP-2 upon ligand stimulation (not shown).

Further evidence identifying overexpressed SIRP α 1 with the endogenous protein in SHP-2 immunoprecipitates from A431 cells was their similar size (compare Fig. 1a, top, lane 6 and Fig. 3a, top) and the fact that they both reacted with monoclonal antibodies raised against a GST fusion protein containing Ex1 in immunoblot analysis (results not shown). Using polyclonal anti-Ex1 antibodies, we immunoprecipitated a protein of apparent M_r 85K–90K from A431 and HBL-100 tumour cells and from human fibroblasts which became tyrosine-phosphorylated upon stimula-

tion with EGF, insulin or PDGF, respectively, and coprecipitated with SHP-2 in a ligand-dependent manner (Fig. 3b). Our data demonstrate the existence of SIRP α subfamily members in several human cell lines which serve as substrates for insulin, EGF and β PDGF receptors, and which in their tyrosine-phosphorylated form bind SHP-2 and may act as its substrates. The interaction of SHP-2 with SIRP α 1 probably involves one or both SH2 domains of the phosphatase, as indicated by the requirement for phosphotyrosine residues (Fig. 3a, c) and the loss of association after mutation of critical residues in SHP-2 SH2 domains (results not shown).

To test for possible association with other SH2-domain proteins, stably overexpressed SIRP α 1 immunoprecipitated from untreated or POV-stimulated 293 cells was incubated with ³⁵S-labelled lysates from 293 cells transiently overexpressing individual SH2-containing proteins. The autoradiograph in Fig. 3c shows these interacting proteins resolved by SDS-PAGE and reveals that SIRP α 1 associates not only with SHP-2 *in vitro* but also with SHP-1 and Grb2 in a stimulation-dependent manner but not with p85, Shc, Grb7, PLC- γ , c-Src, Nck, Vav, GAP or ISGF-3 (results not shown). An inactive SHP-1 mutant has been found to bind an unidentified tyrosine-phosphorylated protein of 90K–95K in human 293 cells¹¹; our results suggest that this protein could be SIRP α 1 or a member of its family.

To investigate the function of SIRP α 1, we generated stable transfectants of NIH3T3 cells expressing SIRP α 1 or mutants carrying either point mutations of the putative SHP-2 tyrosine-binding sites (SIRP α 1-4Y) or a deletion of most of the cytoplasmic region

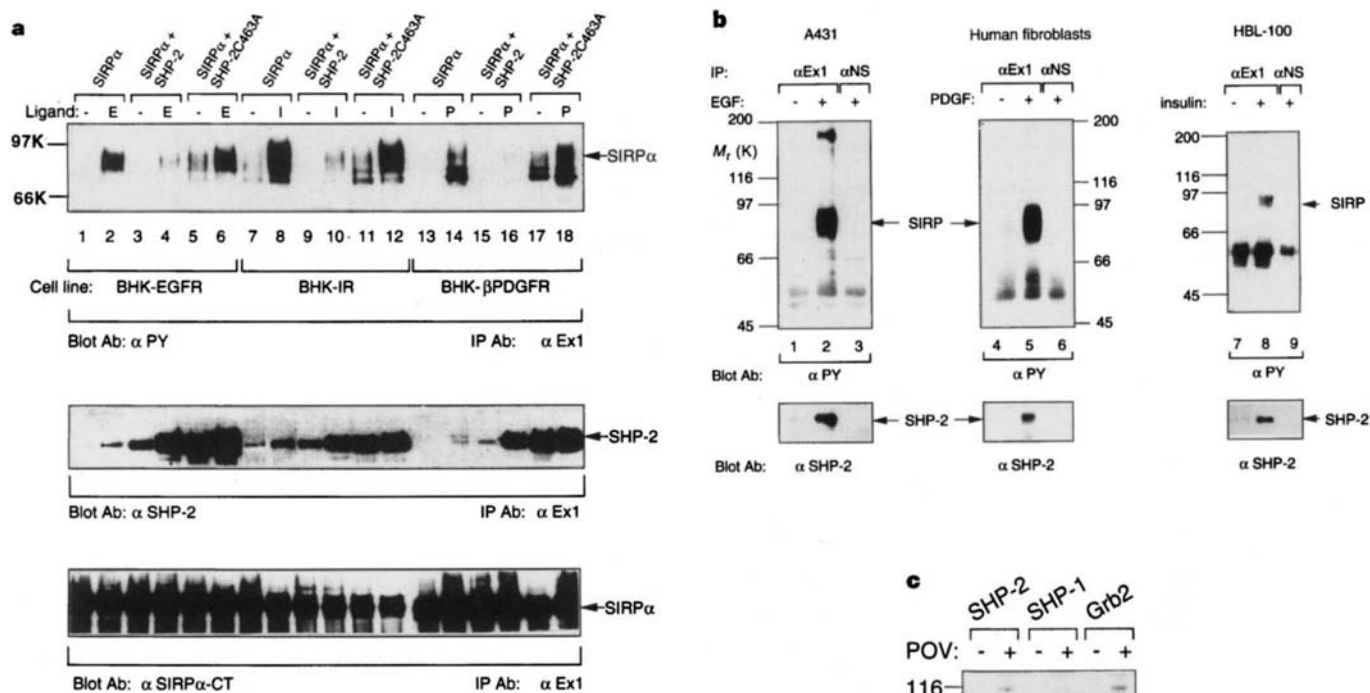


Figure 3 SIRP α 1 binds to SHP-2 and serves as a substrate for IR, EGFR, β PDGFR and SHP-2. **a**, Western blots containing anti-Ex1 immunoprecipitations from quiescent or ligand-stimulated BHK-IR, BHK-EGFR or BHK- β PDGFR cells were labelled with anti-phosphotyrosine (top), anti-SHP-2 (middle) and anti-SIRP α 1-CT antibodies (bottom). **b**, Endogenous SIRP α -like proteins were immunoprecipitated from untreated or EGF-stimulated A431 cells (lanes 1 and 2), from quiescent or PDGF-treated human fibroblasts (lanes 4 and 5), or from starved or insulin-stimulated HBL-100 cells (lanes 7 and 8). As a control, ligand-stimulated cell lysates were immunoprecipitated with preimmune rabbit serum (α NS) (lanes 3, 6 and 9). Immunoblots were probed with monoclonal anti-phosphotyrosine (top) and monoclonal anti-SHP-2 (bottom) antibodies. **c**, *In vitro* association of SIRP α 1 with SHP-2, SHP-1 and Grb2. SIRP α 1-associated ³⁵S-methionine-labelled proteins were resolved on SDS-PAGE and detected by autoradiography (see Methods).

(SIRP α 1- Δ CT). Whereas control cells showed growth-factor-induced DNA synthesis, overexpression of SIRP α 1 led to a significant decrease of 3 H-thymidine incorporation upon stimulation of cells with insulin, EGF (Fig. 4a) or PDGF (results not shown). Both SIRP α 1 mutants had almost no effect on DNA synthesis. As SIRP α 1 became tyrosine-phosphorylated and associated with SHP-2 upon ligand stimulation and SIRP α 1 mutants did not, even

though their expression was comparable (Fig. 4b), the observed inhibition must be connected to SIRP α 1 tyrosine-phosphorylation and/or its association with SHP-2.

The inhibition of growth response upon stimulation with insulin or EGF correlated with reduced activation of MAP kinase in 3T3/SIRP α 1 cells, whereas expression of SIRP α 1 mutants defective in SHP-2 binding had no effect (Fig. 4c). The same result was observed upon stimulation

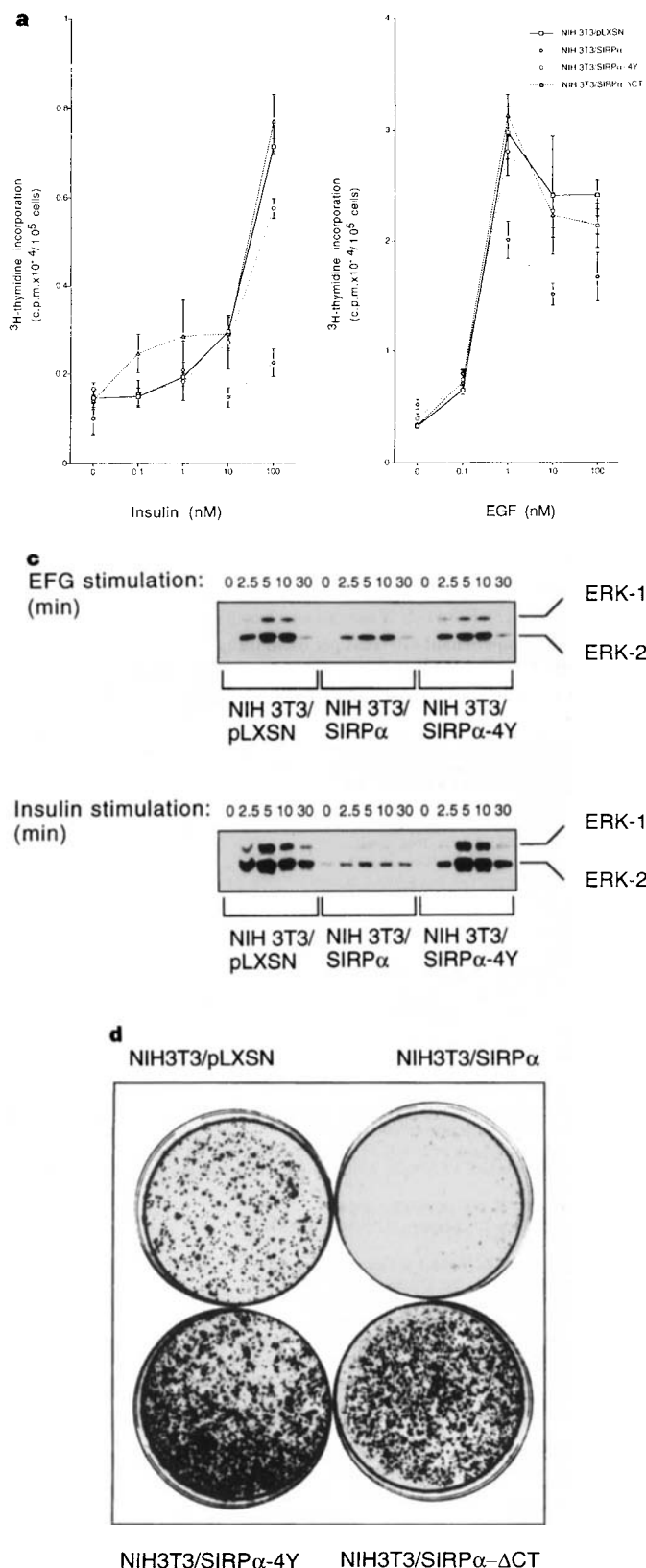


Figure 4 Effects of SIRP α 1 on cell growth and transformation. **a**, Ligand-stimulated 3 H-thymidine incorporation of NIH3T3 cells expressing empty vector (3T3/pLXSN), wild-type SIRP α 1 (3T3/SIRP α 1), SIRP α 1-4Y (3T3/SIRP α 1-4Y) or SIRP α 1- Δ CT (3T3/SIRP α 1- Δ CT; amino acids 402–503 were deleted) mutants. Cells were grown to confluence in 24-well dishes (Nunc), starved for 24 h in DMEM/0.5% FCS, stimulated with different concentrations of insulin or EGF for 18 h, then incubated with 0.5 μ Ci 3 H-thymidine per well for 4 h. Incorporation into DNA was determined as described³⁰. **b**, Western blots showing anti-SIRP α 1 immunoprecipitations from quiescent or ligand-stimulated 3T3/pLXSN, 3T3/SIRP α 1, 3T3/SIRP α 1-4Y or 3T3/SIRP α 1- Δ CT cells were labelled with anti-phosphotyrosine (top), anti-SHP-2 (middle) and anti-Ex1 antibodies (bottom). The expression levels of SIRP α 1, SIRP α 1-4Y and SIRP α 1- Δ CT in NIH3T3 cells were detected in crude cell lysates by immunoblotting using polyclonal anti-Ex1 antibodies. **c**, 3T3/pLXSN, 3T3/SIRP α 1 or 3T3/SIRP α 1-4Y cells were starved for 18 h in DMEM/0.5% FCS and stimulated with insulin or EGF for the time indicated. The activated forms of Erk1 and Erk2 kinases were detected by immunolabelling with a PhosphoPlus MAPK antibody kit (BioLabs). **d**, To examine the ability of SIRP α 1 to influence the formation of cell foci, subconfluent 3T3/pLXSN, 3T3/SIRP α 1, 3T3/SIRP α 1-4Y or 3T3/SIRP α 1- Δ CT cells were infected with *v-fms* virus supernatants.

with PDGF (not shown). These data indicate that SIRP α 1 is a novel regulatory element in the pathway that leads to MAP kinase activation.

We next determined the effect of SIRP α 1 overexpression on oncogene-mediated transformation of NIH3T3 cells. As judged by focus formation, transformation by a *v-fms*-carrying retrovirus was significantly suppressed in cells overexpressing SIRP α 1 but not in 3T3/SIRP α 1-4Y or 3T3/SIRP α 1- Δ CT cells (Fig. 4d).

Overexpression of an inactive SHP-2 mutant was associated with tyrosine hyperphosphorylation of SHP-2-binding proteins of M_r 110K–130K in mouse, rat or hamster cells^{12–14}. Correlating with this, disruption of SHP-2 function induced a variety of negative effects on growth-factor-induced cellular signals^{12,15–21}. Our results indicate that SIRP α 1 and the 90K–130K proteins are members of the same family of regulators.

The mechanism by which SIRP α exerts its negative regulatory function is unknown. As the inhibition by SIRP α 1 of NIH3T3 cell proliferation and transformation depends on phosphorylation of tyrosine residues, these must be important docking sites for SHP-2, SHP-1 or Grb2, or other as-yet unidentified proteins. The dephosphorylation of SIRP α 1 could be tightly regulated by one or both of the SHP phosphatases. Alternatively, or even as well, phosphorylated SIRP α 1 could sequester SHP-2 away from activated RTKs to prevent it exerting other positive regulatory effects, for example by acting as an adapter protein to recruit the Grb2–SOS complex to activated receptors^{22,23}. This idea is supported by the higher affinity of SHP-2 for tyrosine-phosphorylated SIRP α -like proteins rather than for autophosphorylated insulin and EGF receptors^{13,14}. Another possibility is based on the extracellular features of SIRP α 1: the extensive sequence diversity within the immunoglobulin domains, suggests that extracellular determinants participate in specific interactions with receptors, soluble ligands or extracellular matrix components, to influence intracellular signalling.

The characterization of a new family of highly homologous regulators of phosphotyrosine-mediated signals raises questions as to their function and their importance in the pathology of diseases such as cancer and diabetes.

Note added in proof: The structure of another member of the SIRP α family has been reported by Fujioka *et al.*³¹.

Methods

Cell culture and transient expression. MM5/C1, Rat1-IR, A431 or human fibroblast cells were grown until confluency, starved for 18 h in serum-free medium, and either left untreated or were stimulated with POV (1 mM sodium orthovanadate, 3 mM H₂O₂), insulin (100 nM), EGF (1 nM), or PDGF (100 pM) for different time as indicated. SIRP α 1, SHP-2 (ref. 24) or SHP-2C463A mutant¹⁰ cDNAs were transiently cotransfected in BHK-IR, BHK-EGFR or BHK- β PDGFR cells using calcium phosphate precipitation²⁵. After stimulation, cells were lysed in buffer containing 50 mM HEPES, pH 7.5, 150 mM NaCl, 1% Triton X-100, 10% glycerol, 1 mM POV, 1 mM EDTA, 1 mM PMSF, 1 μ g ml⁻¹ leupeptin, 1 μ g ml⁻¹ aprotinin.

Immunoprecipitation and western blotting. SHP-2 immunoprecipitations were performed with polyclonal anti-SHP-2 antibodies²⁴. Overexpressed SIRP α 1 or endogenous SIRP α 1-like proteins were immunoprecipitated by polyclonal anti-Ex1 antibodies raised by immunizing rabbits with a GST fusion protein containing the Ex1 fragment (Fig. 2b). Western blots were labelled with monoclonal anti-phosphotyrosine antibody 5E2 (ref. 26), and after stripping, reprobed with monoclonal anti-SHP-2 antibodies (Transduction Labs), or polyclonal anti-SIRP α 1-CT antibodies, raised against a GST fusion protein containing the C-terminal part of SIRP α 1 (amino acids 336–503). For immunolabelling, goat anti-mouse or anti-rabbit horseradish peroxidase conjugates (Bio-Rad) and the ECL detection system (Amersham) were used.

For the *in vitro* association experiment, SIRP α 1 was immunoprecipitated with polyclonal anti-Ex1 antibodies from quiescent or POV-stimulated 293 cells that stably overexpressed SIRP α 1. Subsequently, crude lysates of ³⁵S-methionine-labelled 293 cells overexpressing different SH2-domain proteins were added to the affinity matrix and incubated for 2 h at 4 °C. Immunocomplexes were washed, separated by SDS–PAGE and analysed by autoradiography.

Enzymatic deglycosylation *in vitro*. SHP-2 immunocomplexes were first denatured in 1% SDS at 100 °C for 5 min. Deglycosylation was in potassium phosphate buffer (40 mM, pH 7.0), containing 20 mM EDTA, 1% β -mercaptoethanol, 1% Triton X-100 and 0.5 units of Endoglycosidase F/N-Glycosidase F (Boehringer Mannheim) at 37 °C for 16 h.

Protein purification. About 10¹⁰ Rat1-IR cells were used to purify the 110K protein. Starved Rat1-IR cells were insulin-stimulated (100 nM) for 10 min, washed briefly with ice-cold hypotonic buffer containing 20 mM HEPES, pH 7.5, 1 mM POV, 1 mM EDTA, 1 mM PMSE, 1 μ g ml⁻¹ leupeptin, 1 μ g ml⁻¹ aprotinin, scraped into the same buffer and homogenized. Cell extracts were pelleted at 1,000 r.p.m. for 15 min, and supernatants were spun at 48,000g for 1 h. Membranes were solubilized in lysis buffer. hIR was depleted from membrane extracts using an affinity column with monoclonal anti-hIR antibody 83–14 (ref. 27) covalently coupled to protein A–Sepharose beads (Pharmacia). Depleted extracts were applied onto a WGA–agarose 6MB column (Sigma), and glycoproteins were eluted with 0.3 M *N*-acetyl glucosamine in HNTG (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.1% Triton X-100, 10% glycerol, 1 mM POV). After concentration, protein extracts were applied to an anti-phosphotyrosine antibody column (Sigma). Bound proteins were eluted with 20 mM phosphotyrosine in HNTG. The eluate was resolved by SDS–PAGE, proteins were transferred to a PVDF membrane (Millipore) and stained with Coomassie blue. PVDF membrane slices corresponding to the 110K position were microsequenced.

Biological assays. To produce retroviruses expressing pLXSN²⁸, and wild-type SIRP α 1 and mutant SIRP α 1 constructs, BOSC23 cells were transiently transfected by expression plasmids as described²⁹. To obtain NIH3T3 cells stably expressing wild-type SIRP α 1, SIRP α 1-4Y or SIRP α 1- Δ CT mutants, subconfluent NIH3T3 cells (10⁵ cells per 6-cm dish) were incubated with supernatants of transfected BOSC 23 cells for 4 h in the presence of Polybrene (4 μ g ml⁻¹), followed by selection with G418 (1 mg ml⁻¹).

For focus-formation assays, cell lines 3T3/pLXSN, 3T3/SIRP α 1, 3T3/SIRP α 1-4Y or 3T3/SIRP α 1- Δ CT were superinfected for 4 h with equal volumes of *v-fms* virus supernatant (10⁵ cells per 6-cm dish). Cells were cultivated for 14 d in 4% FCS with medium change every second day. Cell foci were stained with crystal violet (0.1% crystal violet, 30% methanol).

Received 28 August 1996; accepted 21 January 1997.

- Barford, D., Jia, Z. & Tonks, N. K. *Nature Struct. Biol.* **2**, 1043–1053 (1995).
- Traverse, S. *et al. Curr. Biol.* **4**, 694–701 (1994).
- Dikic, I., Schlessinger, J. & Lax, I. *Curr. Biol.* **4**, 702–708 (1994).
- Obermeier, A., Tinhofer, I., Grunick, H. H. & Ullrich, A. *EMBO J.* **15**, 73–82 (1996).
- Saltiel, A. R. *FASEB J.* **8**, 1034–1040 (1994).
- Matozaki, T. & Kasuga, M. *Cell. Signal.* **8**, 13–19 (1996).
- Streuli, M. *Curr. Opin. Cell Biol.* **8**, 182–188 (1996).
- Holgado-Madruga, M., Emler, D. R., Moscatello, D. K., Godwin, A. K. & Wong, A. J. *Nature* **379**, 560–564 (1996).
- Herbst, R. *et al. Cell* **85**, 899–909 (1996).
- Stein-Gerlach, M., Kharitonov, A., Vogel, W., Ali, S. & Ullrich, A. *J. Biol. Chem.* **270**, 24635–24637 (1995).
- Su, L. *et al. J. Biol. Chem.* **271**, 10385–10390 (1996).
- Milarski, K. L. & Saltiel, A. R. *J. Biol. Chem.* **269**, 21239–21243 (1994).
- Yamauchi, K., Ribons, V., Saltiel, A. R. & Pessin, J. E. *J. Biol. Chem.* **270**, 17716–17722 (1995).
- Yamauchi, K. & Pessin, J. E. *J. Biol. Chem.* **270**, 14871–14874 (1995).
- Yamauchi, K., Milarski, K. L., Saltiel, A. R. & Pessin, J. E. *Proc. Natl Acad. Sci. USA* **92**, 664–668 (1995).
- Sawada, T., Milarski, K. L. & Saltiel, A. R. *Biochem. Biophys. Res. Commun.* **214**, 737–743 (1995).
- Xiao, S. *et al. J. Biol. Chem.* **269**, 21244–21248 (1994).
- Noguchi, T., Matozaki, T., Horita, K., Fujioka, Y. & Kasuga, M. *Mol. Cell. Biol.* **14**, 6674–6682 (1994).
- Zho, Z. *et al. J. Biol. Chem.* **270**, 11765–11769 (1995).
- Tang, T. L., Freeman, R. M. Jr, O'Reilly, A. M., Neel, B. G. & Sokol, S. Y. *Cell* **80**, 473–483 (1995).
- Bennett, A. M., Hausdorff, S. F., O'Reilly, A. M., Freeman, R. M. & Neel, B. G. *Mol. Cell. Biol.* **16**, 1189–1202 (1996).
- Li, W. *et al. Mol. Cell. Biol.* **14**, 509–517 (1994).
- Bennett, A. M., Tang, T. L., Sugimoto, S., Walsh, C. T. & Neel, B. G. *Proc. Natl Acad. Sci. USA* **9**, 7335–7339 (1994).
- Vogel, W., Lammers, R., Huang, J. & Ullrich, A. *Science* **259**, 1611–1614 (1994).
- Chen, C. & Okayama, H. *Mol. Cell. Biol.* **7**, 2745–2752 (1987).
- Fendly, B. M. *et al. Cancer Res.* **50**, 1550–1558 (1990).
- Soos, M. A. *et al. Biochem. J.* **235**, 199–208 (1986).
- Miller, A. D. & Rosman, G. J. *Biotechniques* **7**, 980–988 (1989).
- Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. *Proc. Natl Acad. Sci.* **90**, 8392–8396 (1993).
- Redemann, N., Holzmann, B., Wagner, E. F., Schlessinger, J. & Ullrich, A. *Mol. Cell. Biol.* **12**, 491–498 (1992).
- Fujioka, Y. *et al. Mol. Cell. Biol.* **16**, 6887–6899 (1996).

Acknowledgements. We thank W. Vogel for participation during the initial protein purification, G. Plowman for database searches, R. Lammers for *v-fms* virus supernatants, S. Werner for the human fibroblast cell line, and S. Gartner for technical assistance. This work was supported by a grant from SUGEN, Inc (Redwood City).

Correspondence and requests for materials should be addressed to A.U. (e-mail: ullrich@biochem.mpg.de).

Phosphorylation of RNA-binding protein controls cell cycle switch from mitotic to meiotic in fission yeast

Yoshinori Watanabe[†], Satoko Shinozaki-Yabana, Yuji Chikashige[†], Yasushi Hiraoka[†] & Masayuki Yamamoto

Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Tokyo 113, Japan

[†] Kansai Advanced Research Center, Communications Research Laboratory, 588-2 Iwaoka, Iwaoka-cho, Nishi-ku, Kobe 651-24, Japan

Meiosis generates haploid gametes from diploid cells and is an almost universal feature of eukaryotic organisms. But little is known about how the switch from mitotic to meiotic cell cycles is molecularly controlled. In the fission yeast *Schizosaccharomyces pombe*, inactivation of the protein kinase Pat1(Ran1) upon nutritional deprivation triggers entry into the meiotic cell cycle¹⁻⁶. Here we show that the RNA-binding protein Mei2 is a substrate of Pat1 kinase and that dephosphorylation of Mei2 is sufficient to switch cells from the mitotic cell cycle into meiosis. Mei2 is localized mainly in the cytoplasm of proliferating cells but is seen as a single spot close to the microtubule organizing centre in prophase nuclei during meiosis. Our results, and others from a metazoan⁷, emphasize the crucial role of RNA-binding proteins in the initiation and execution of meiosis.

The onset of meiosis in *S. pombe* is regulated by the nutritional conditions and the mating type of the cell^{6,8}. Nutritional deprivation induces expression of *mei3*⁺ in diploid cells heterozygous for the mating type, and Mei3 protein inhibits the activity of Pat1 kinase^{4,5}. The *mei2* gene is indispensable for meiosis⁸. The null mutation in *mei2* can suppress *pat1*-driven haploid meiosis^{3,9,10}, suggesting that Mei2 may act downstream of Pat1 kinase. Mei2 is an RNA-binding protein carrying three potential RNA-recognition motifs^{11,12} (Fig. 1a), structurally similar to *Drosophila* ELAV¹³ and related proteins. It is required for both premeiotic DNA synthesis^{3,8} and the first meiotic division, and cooperates with a unique polyadenylated RNA molecule called meiRNA to promote the first division¹².

Deletion analysis of Mei2 unexpectedly showed that removal of the amino-terminal two RRM does not impair its function seriously (Fig. 1a), although a point mutation in the N-terminal RRM made the protein heat-sensitive¹². In contrast, either removal of the carboxy-terminal 38 amino acids or a point mutation in the C-terminal RRM (F644A)¹² inactivated its function completely. A Mei2 derivative consisting of residues 429–733, containing only the C-terminal RRM, was the shortest construct to be functional *in vivo* (Fig. 1a). A glutathione-S-transferase (GST) fusion protein carrying residues 429–750 (GST–Mei2C) could bind to meiRNA *in vitro*, as did another carrying residues 30–750 (GST–Mei2F) (Fig. 1b).

We expressed the entire *mei2*⁺ ORF from the thiamine-repressible *nmt1* promoter in *mei2Δpat1*⁺ and *mei2Δpat1Δ* haploid strains. When transcription of *mei2*⁺ was induced by removal of thiamine, the *pat1Δ* strain halted vegetative growth and initiated haploid meiosis, whereas the *pat1*⁺ strain continued growth (Fig. 1c). These results indicate that Pat1 kinase can block *mei2*⁺ function post-transcriptionally. Furthermore, yeast two-hybrid analysis suggested that Mei2 interacted with the kinase domain of Pat1 (data not shown). We therefore did an *in vitro* phosphorylation assay using

recombinant GST–Pat1 and GST–Mei2 fusion proteins. GST–Pat1 efficiently phosphorylated both GST–Mei2F and GST–Mei2C, but barely phosphorylated GST–Mei2N carrying the N-terminal region (residues 30–428) (Fig. 2a). The addition of GST–Mei3, which inhibits Pat1 kinase⁴, blocked the phosphorylation quantitatively (Fig. 2b). We conclude that Pat1 can phosphorylate Mei2 *in vitro*, predominantly in its C-terminal region.

To investigate this phosphorylation of Mei2 *in vivo*, we labelled non-functional Mei2–F644A metabolically with ³²P in *pat1*⁺ and *pat1Δ* cells, and recovered it by immunoprecipitation. We prepared *in vitro*-phosphorylated GST–Mei2C as a reference. Phosphoproteins were subjected to tryptic phosphopeptide mapping (Fig. 2c). GST–Mei2C gave three major phosphopeptides, A–C. The Mei2 protein recovered from *pat1*⁺ cells showed a similar pattern, except for phosphopeptide D, whose phosphorylation apparently does not depend on Pat1. Phosphopeptides A to C were not detected in *pat1Δ* cells. We confirmed the identity of the corresponding peptides by analysis of a mixture of the two samples (data not shown). These results indicated that Pat1 kinase phosphorylates Mei2 *in vivo* as it does *in vitro*.

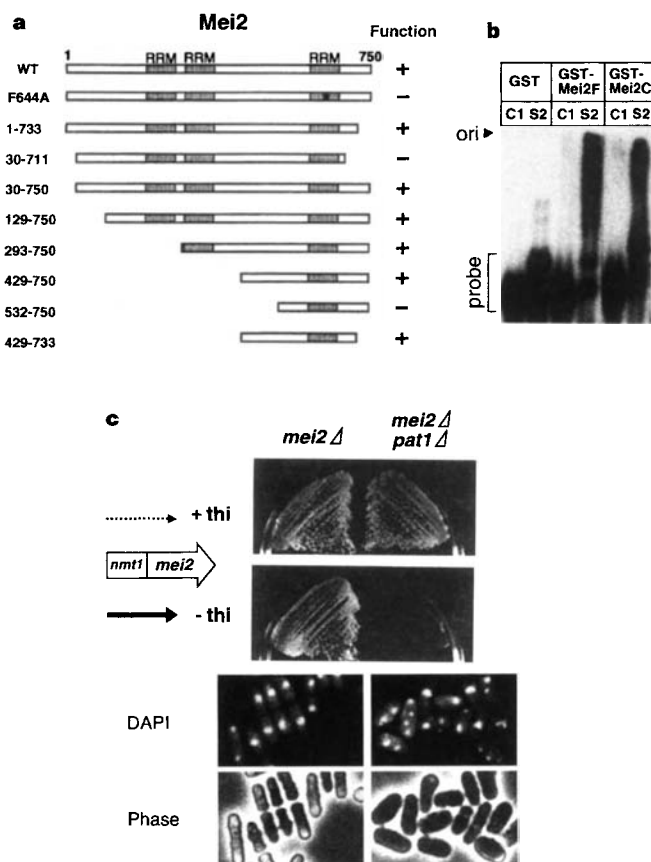


Figure 1 Delimitation of the functional domain of Mei2 and its genetic interaction with Pat1. **a**, Deletion derivatives of Mei2 and their ability to provide *mei2* function. Hatched boxes indicate RNA-recognition motifs (RRMs). Each derivative was expressed from the *nmt1* promoter on pREP1 in diploid *mei2Δ* strain. Transformants were tested for their ability to sporulate after 3 days incubation on synthetic sporulation agar (SSA) at 30°C. **b**, Binding of the C-terminal half of Mei2 to meiRNA. Recombinant GST, GST–Mei2F(30–750) and GST–Mei2C(429–750) were mixed with control RNA (C1) and meiRNA (S2) as described¹². **c**, *pat1*⁺ blocks function of *mei2*⁺ at the post-transcriptional level. We streaked *mei2Δpat1*⁺ and *mei2Δpat1Δ* haploid cells, transformed with pREP1–*mei2*⁺, onto an SSA plate either with or without thiamine and incubated them at 30°C for 3 d (top). We examined the same cells, cultured in liquid MM without thiamine for 20 h, by DAPI staining and phase-contrast microscopy (bottom).

* Present address: ICRF Cell Cycle Laboratory, Lincoln's Inn Fields, London WC2A 3PX, UK

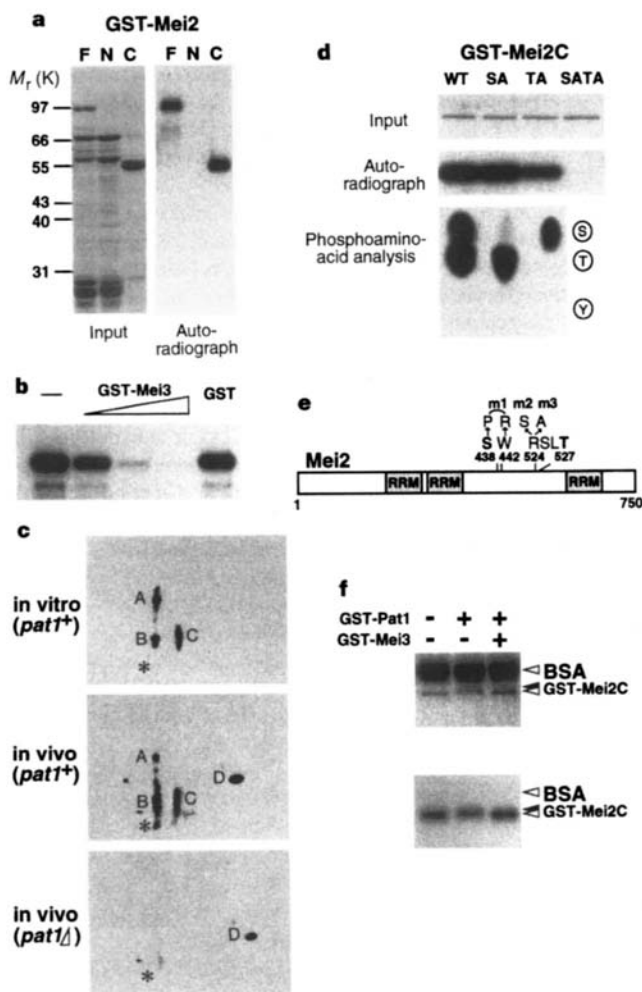


Figure 2 Phosphorylation of Mei2 by Pat1. **a**, Phosphorylation of Mei2 *in vitro*. Recombinant GST-Mei2 carrying Mei2 residues 30–750 (F), 30–428 (N) or 429–750 (C) were resolved by SDS-PAGE (11%) (left; Coomassie blue staining). The top band in each lane appeared to represent the respective fusion protein, and others are likely to be its degradation products. Proteins were incubated with GST-Pat1 in K buffer containing [α - 32 P]ATP and separated by gel electrophoresis. The gel was dried and autoradiographed (right panel). **b**, Mei3 blocks Pat1-dependent phosphorylation of Mei2 *in vitro*. GST-Mei2C was phosphorylated in the presence of purified GST-Mei3. GST-Mei3 was added in 1 $\times$, 2 $\times$ or 8 $\times$ molar excess to GST-Pat1. GST, as a control, was added in 8 $\times$ excess. **c**, Tryptic phosphopeptide mapping of *in vitro* and *in vivo* phosphorylated Mei2. GST-Mei2C phosphorylated *in vitro* by GST-Pat1 with [α - 32 P]ATP was subjected to two-dimensional tryptic phosphopeptide analysis (top panel). Mei2 protein precipitated either from 32 P-labelled *pat1*⁺ cells (middle panel) or from *pat1* Δ cells (bottom panel) was also analysed. The principal phosphopeptides are marked as A–D. Asterisks indicate the position of sample application. **d**, Pat1 phosphorylates one serine and one threonine residue in Mei2. GST-Mei2C and its mutant versions were phosphorylated by GST-Pat1 *in vitro* and analysed by SDS-PAGE. The input proteins (top panel) were autoradiographed after phosphorylation (middle panel). Each 32 P-labelled GST-Mei2C variant was subjected to one-dimensional phosphoaminoacid analysis (bottom panel). S, phosphoserine; T, phosphothreonine; Y, phosphotyrosine. **e**, Amino-acid alterations in 'activated' Mei2 (m1–m3); the two assigned phosphorylation sites of Mei2 are shown in bold. **f**, Ultraviolet crosslinking assay of binding of Mei2 to meiRNA. GST-Mei2C was treated with GST-Pat1 in the presence or absence of GST-Mei3. Each reaction mix was then incubated with 32 P-labelled meiRNA and irradiated with ultraviolet light. After digestion with RNase A, products were analysed by SDS-PAGE and coomassie blue staining (upper panel) followed by autoradiography (lower panel). GST-Mei2C (middle lane) ran more slowly than the other two as it was phosphorylated.

Two lines of analysis facilitated identification of the phosphorylated sites of Mei2 by Pat1. One is *in vitro* phosphorylation of deletion derivatives of GST-Mei2C, and the other is isolation of mutant *mei2* alleles that gave constitutively active gene products. The former analysis suggested that a serine residue between positions 429–475 and a threonine residue between residues 502–531 are likely to be phosphorylated in Mei2 (data not shown). The latter analysis identified three 'activated' *mei2* alleles (Fig. 2e). The first allele (m1) carried two mutations: Ser 438 to proline and Trp 442 to arginine. Arg 524 was replaced in the second (m2) and third (m3) alleles, by alanine and serine respectively (Fig. 2e). These results indicated that Pat1 probably phosphorylates Ser 438 and Thr 527. Accordingly, we prepared three types of mutant GST-Mei2C, in which (1) Ser 438 was replaced by alanine (denoted as SA), (2) Thr 527 was replaced by alanine (TA), and (3) both residues were replaced by alanine (SATA). These and control GST-Mei2C were phosphorylated by Pat1 *in vitro* and analysed for phosphoaminoacids (Fig. 2d). Wild-type GST-Mei2C generated both phosphoserine and phosphothreonine, but phosphoserine was missing in the SA mutant and phosphothreonine was missing in the TA mutant. No significant phosphorylation was detected for the SATA mutant. As the N-terminal half of Mei2 does not appear to be phosphorylated significantly by Pat1 (Fig. 2a), these results indicate that Pat1 phosphorylates Mei2 on Ser 438 and Thr 527 almost exclusively. Arginine precedes these residues, with two amino acids in between (RxxS/T), like target sites for many protein-serine/threonine kinases. Phosphopeptide analysis of the mutant proteins revealed that phosphopeptide A (Fig. 2c) carries phosphothreonine, whereas both B and C carry phosphoserine (data not shown).

Both Ser 438 and Thr 527 are located in the auxiliary domain between the second and third RNA-recognition motifs (Fig. 2e), which is indispensable for the Mei2 function (Fig. 1a). Ultraviolet crosslinking experiments indicated that phosphorylation does not significantly affect the RNA-binding ability of Mei2 (Fig. 2f). Artificial expression of non-phosphorylatable Mei2 gave striking results, however: as shown in Fig. 3a, Mei2-SATA expressed from the weak *nmt1* promoter blocked cell growth and induced ectopic meiosis and sporulation in haploid cells without starvation. Wild-type Mei2, Mei2-SA and Mei2-TA did not show significant effects under the same conditions.

We followed meiosis caused by the expression of *mei2*-SATA integrated in the chromosome. Strains carrying either *nmt1*-*mei2*-SATA or *nmt1*-*mei2*⁺ were shifted to thiamine-free minimal medium. Transcripts of *mei2* started to accumulate between 12 and 14 hours after the shift in both cases, reaching a level comparable to that of physiologically induced meiosis (Fig. 3b). Cells carrying *nmt1*-*mei2*-SATA began to exhibit a G1 population concomitantly with expression of *mei2*, but cells carrying *nmt1*-*mei2*⁺ did not (Fig. 3c). Nearly all the *mei2*-SATA cells had one nucleus at 16 h after the shift (Fig. 3d). The cell number reached a plateau around this point and the cells displayed few septa thereafter (data not shown). These observations and the flow cytometric data (Fig. 3c) suggest that the *mei2*-SATA population at 16 h was mainly composed of cells in G1 and meiotic prophase. The number of cells with two nuclei then increased transiently, followed by cells with three or four nuclei (Fig. 3d). At 20 h, nearly 50% of *mei2*-SATA cells had 3–4 nuclei (Fig. 3d, e). The peak formed by cells with two nuclei (Fig. 3d) implies that progression of the first meiotic division was fairly synchronous. Transcription of the meiosis-specific gene *spo5* was fully induced during this process, reaching a peak at 18 h (Fig. 3b). Transcription of the *rep2* gene, which encodes a transcriptional activator subunit of a G1/S-specific transcription factor¹⁴, was obviously reduced in response to the expression of *mei2*-SATA (Fig. 3b). As reduction of *rep2*⁺ expression in growing cells leads to a decrease in expression of *cdc22*, which encodes ribonucleotide reductase¹⁵, and to the accumulation of cells in G1 (ref. 14), this may explain why Mei2-SATA causes cell-cycle arrest in G1. About 30% of

mei2-SATA cells remained arrested at G1 and did not initiate meiosis in this system. The reason for this is unclear, but it suggests that the ability of Mei2 to arrest the mitotic cell cycle at G1 is not necessarily linked to the ability to induce premeiotic DNA synthesis. Transcription of *rep1*⁺, a putative meiotic counterpart of *rep2*⁺ (ref. 16), was induced by the expression of *mei2-SATA* (Fig. 3b; Y. Iino and M.Y., unpublished results). The transcriptional switch between *rep1*⁺ and *rep2*⁺ was not observed in the *mei2*⁺ cells, indicating that these genes are controlled by Mei2. It is notable that Mei2, a meiotic regulator dispensable for mitotic growth, can influence the transcription of a gene controlling the mitotic cell cycle. In contrast to these dramatic changes in *mei2-SATA* cells, control *mei2*⁺ cells continued mitotic growth (Fig. 3c, d), although Mei2 protein had accumulated in them (data not shown). These results show that dephosphorylated Mei2 protein can arrest the mitotic cell cycle and induce meiotic differentiation.

We investigated the subcellular localization of Mei2 protein by using the peptide-tag method with jellyfish green fluorescent protein (GFP)¹⁷. A *mei2*⁺ ORF–GFP fusion, which could provide *mei2*⁺ function, was expressed from the weak *nmt1* promoter. Mei2–GFP was distributed mainly in the cytoplasm in exponentially growing cells (data not shown). Curiously, however, Mei2–GFP moved to the nucleus and formed a clear single dot when the

cells proceeded to meiotic prophase (Fig. 4). Observation of the Mei2 dot in moving ‘horse-tail’ prophase nuclei indicated that it occupied a fixed position in the nucleus, which was separate but not far from the microtubule organizing centre and the telomere cluster, which are located at the leading tip of horse-tail nuclei¹⁸ (data not shown). In cells of several vertebrates and plants, including *Xenopus* eggs, a conspicuous organelle called the ‘coiled body’ has been described, which contains ribonucleoprotein complexes and is thought to be involved in RNA processing¹⁹. The relation between the Mei2 dot and the coiled body remains to be elucidated.

It has been shown that *in vitro* Pat1 can phosphorylate Ste11 (ref. 20), a key transcription factor that regulates developmental genes including *mei2* (ref. 21). However, mutations in the phosphorylation sites of Ste11 caused no obvious meiotic phenotype²⁰. We confirmed that weak expression of *mei2-SATA* induces meiosis even in *ste11*-deleted cells. Thus, although dephosphorylation of Ste11 may increase the readiness for meiosis by enhancing expression of *mei2*, the most critical control exerted by Pat1 kinase over meiotic entry is phosphorylation of Mei2. □

Methods

Mating types of the strains. We used *h*[−] haploid and *h*⁺/*h*[−] diploid strains throughout this study.

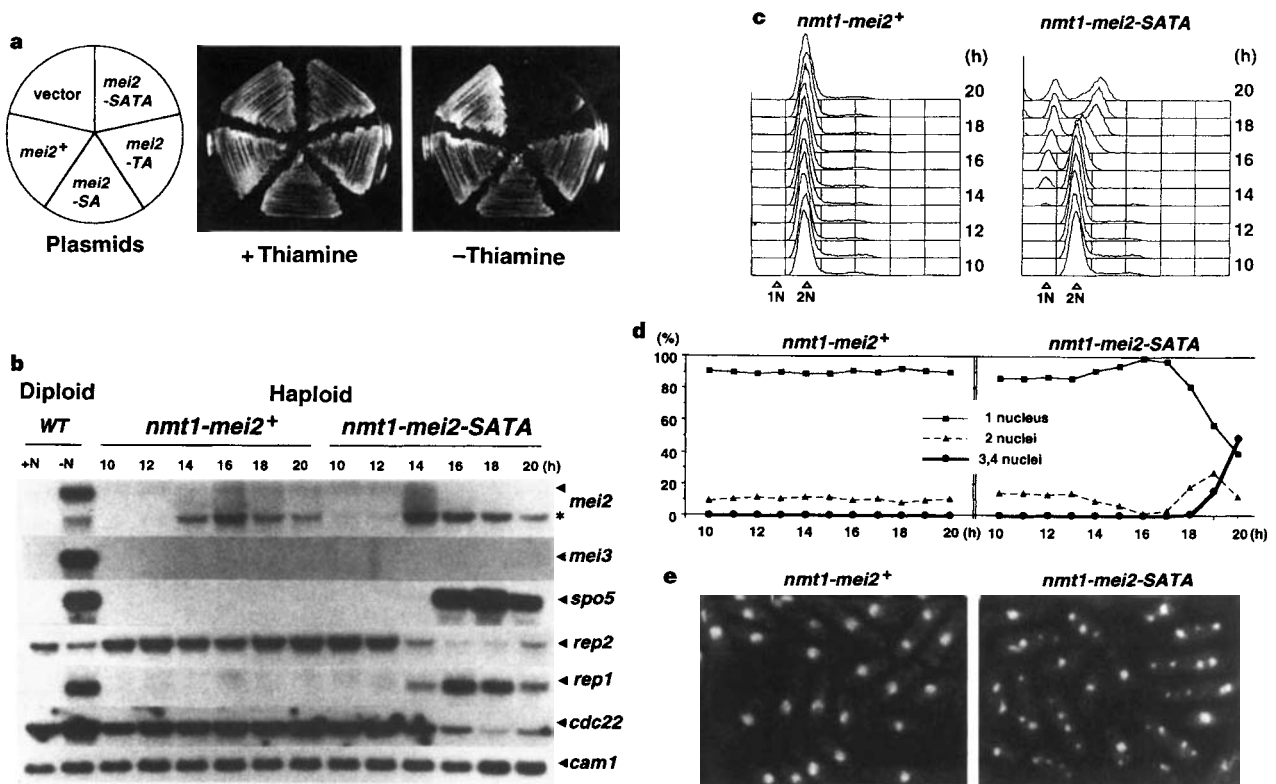


Figure 3 Meiosis induced by Mei2-SATA. **a**, A wild-type haploid strain was transformed with pREP81 carrying either *mei2*⁺, *mei2-SA*, *mei2-TA*, *mei2-SATA* or vector alone. Each transformant, recovered from a plate containing thiamine, was restreaked on SSA plates with and without thiamine and incubated at 30°C for 3 days. **b–e**, The *nmt1-mei2*⁺ and *nmt1-mei2-SATA* haploid integrants were cultured in liquid MM with no thiamine for the periods indicated. Cells in each sample were collected and analysed by northern blotting (**b**), flow cytometry (**c**), and DAPI staining (**d** and **e**). The cell concentration was kept between $5 \times 10^6 \text{ ml}^{-1}$ and $8 \times 10^6 \text{ ml}^{-1}$ by supplying fresh medium hourly during this analysis (10–20 h). **b**, RNA was prepared from each sample and tested for expression of the indicated genes. RNA from wild-type diploid cells, either growing exponentially (+N) or

nitrogen-starved for 4 h (−N), was used as a reference. Asterisk indicates *mei2* transcripts read from the *nmt1* promoter; arrowhead indicates endogenous *mei2*⁺ mRNA, expression of which is enhanced by nitrogen starvation²⁸. Features of genes analysed are as follows: transcription of *mei3* and *spo5* is meiosis-specific^{4,12}, transcription of *rep2* and *rep1* is related to cell-cycle control^{14,16}, and expression of *cdc22* is essential for DNA synthesis⁵. Transcripts of *cam1* encoding calmodulin²⁹ are shown as a loading control. **c**, Flow cytometry to determine the cellular DNA content using a Becton–Dickinson FACScan apparatus. Growing cells have a G2-equivalent DNA content (2N) in this organism. **d**, The number of nuclei in a cell counted after DAPI staining. **e**, Nuclei in cells cultured for 20 h in the absence of thiamine and stained with DAPI.

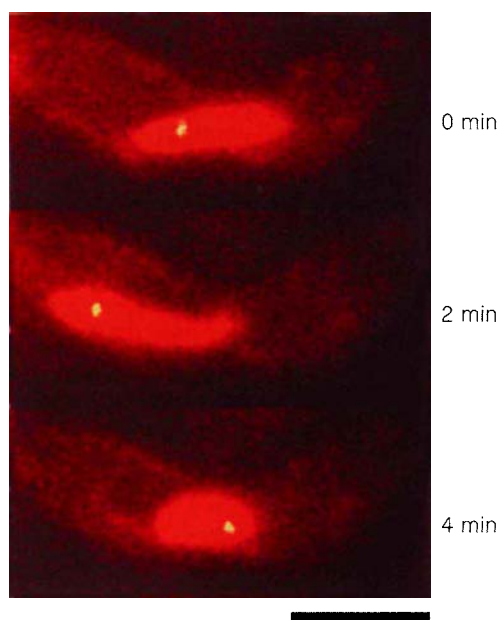


Figure 4 Live video observation of the Mei2-GFP dot in meiotic prophase. A diploid cell generated by conjugation was monitored. Three images taken at 2-min intervals are shown. Red regions, nuclear DNA; yellow spots, fluorescence of Mei2-GFP. The whole-cell body is faintly visible in red. Scale bar, 5 μ m.

Expression vectors. pREP1 carries the authentic thiamine-repressible *nmt1* promoter, whereas pREP41 and pREP81 carry its modified versions, the expression of which is respectively 10-fold and 100-fold lower than that of pREP1 (ref. 22). To express full-length *mei2* from the *nmt1* promoter, an *NdeI* site was created at the initiation codon of the *mei2* open reading frame (ORF) and a 2.6-kb *NdeI*–*BglII* fragment was cloned into pREP vectors. Integration of *nmt1*–*mei2*⁺ or *nmt1*–*mei2*–SATA on pREP41 into the chromosomal *leu1* locus was done, as described²³. We used minimal medium SD with 1% glucose, which contains thiamine (1.2 μ M), to repress the *nmt1* promoter, and minimal medium MM with 1% glucose to derepress it. Cell growth and sporulation were monitored on synthetic agar plates SSA with or without 2 μ M thiamine (see ref. 12 for more details of the media).

In vitro kinase assay and tryptic phosphopeptide mapping. Full-length Pat1²⁴ and full-length Mei3⁴, each fused to GST, were used in the *in vitro* phosphorylation assay. The assay was carried out as follows: 3 pmol of GST–Mei2 was mixed with 0.2 pmol GST–Pat1 and 2 μ g BSA in K buffer (150 mM Tris–HCl, pH 7.5, 20 mM MgCl₂, 10 mM MgCl₂, 15 mM 2-mercaptoethanol). The reaction was initiated by adding 20 μ M [γ -³²P]ATP and terminated by adding SDS–PAGE buffer after 30 min incubation at 30 °C. To label Mei2 protein *in vivo*, we transformed *mei2* Δ pat1⁺ and *mei2* Δ pat1 Δ haploid strains with pREP41–*mei2*–FA, carrying the inactive F644A allele¹² (Fig. 1a). *mei2*–FA, rather than *mei2*⁺, was used here because viable transformants could be obtained from a *pat1* Δ strain. Transformed cells were labelled with ortho-[³²P]-phosphate for 18 h in MM liquid, collected by centrifugation, and broken in 10% TCA with glass beads. After washing with acetone, the pellet was dissolved in S buffer (50 mM Tris–HCl, pH 7.5, 1 mM EDTA, 1 mM 2-mercaptoethanol, 0.5 mM PMSE, 1% SDS). An aliquot of the supernatant was immunoprecipitated with anti-Mei2 antibodies¹². The immunocomplex was resolved by SDS–PAGE and the Mei2 band eluted from the gel and digested with trypsin as described²⁵. GST–Mei2C phosphorylated *in vitro* was processed similarly. Tryptic samples were applied to thin-layer cellulose plates and separated by high-voltage electrophoresis (at 1,000 V) at pH 1.9 for 30 min, followed by ascending chromatography.

Screening for *mei2* alleles encoding constitutively active gene products. We randomly mutagenized the C-terminal half of *mei2* encoding residues 429–750 by polymerase chain reaction (PCR) and expressed their full-length versions from pREP41 in wild-type haploid cells. Three relatively weak

‘activated’ *mei2* alleles, which could induce haploid meiosis upon induction of transcription, were isolated and sequenced.

Fluorescence microscopy of GFP-tagged Mei2 protein. We cloned a mutant version (Ser65 $\rightarrow$ Thr) of GFP²⁶ into pREP81, and inserted an *NdeI*–*Sall* fragment carrying the exact *mei2* ORF before the GFP ORF. The fusion protein Mei2–GFP was expressed in *mei2* Δ cells to eliminate the competitive effects of endogenous Mei2 protein. Living cells were counterstained with Hoechst 33342 and video-recorded by a Peltier-cooled CCD camera (Photometrics) attached to an Olympus inverted microscope IMT-2, as described^{18,27}.

Received 4 November 1996; accepted 13 January 1997.

1. Iino, Y. & Yamamoto, M. *Mol. Gen. Genet.* **198**, 416–421 (1985).
2. Nurse, P. *Mol. Gen. Genet.* **198**, 497–502 (1985).
3. Beach, D., Rodgers, L. & Gould, J. *Curr. Genet.* **10**, 297–311 (1985).
4. McLeod, M., Stein, M. & Beach, D. *EMBO J.* **6**, 729–736 (1987).
5. McLeod, M. & Beach, D. *Nature* **332**, 509–514 (1988).
6. Yamamoto, M. *Trends Biochem. Sci.* **21**, 18–22 (1996).
7. Eberhart, C. G., Maines, J. Z. & Wasserman, S. A. *Nature* **381**, 783–785 (1996).
8. Egel, R. in *Molecular Biology of the Fission Yeast* (eds Nasim, A. et al.) 31–73 (Academic, San Diego, 1989).
9. Iino, Y. & Yamamoto, M. *Proc. Natl Acad. Sci. USA* **82**, 2447–2451 (1985).
10. Watanabe, Y., Iino, Y., Furuhata, K., Shimoda, C. & Yamamoto, M. *EMBO J.* **7**, 761–767 (1988).
11. Birney, E., Kumar, S. & Krainer, A. R. *Nucleic Acids Res.* **12**, 5803–5816 (1993).
12. Watanabe, Y. & Yamamoto, M. *Cell* **78**, 487–498 (1994).
13. Robinow, S., Campos, A., Yao, K. & White, K. *Science* **242**, 1570–1572 (1988).
14. Nakashima, N., Tanaka, K., Sturm, S. & Okayama, H. *EMBO J.* **14**, 4794–4802 (1995).
15. Fernandez-Sarabia, M. J., McInerney, C., Harris, P., Gordon, C. & Fantes, P. *Mol. Gen. Genet.* **238**, 241–251 (1993).
16. Sugiyama, A., Tanaka, K., Okazaki, K., Nojima, H. & Okayama, H. *EMBO J.* **13**, 1881–1887 (1994).
17. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. *Science* **263**, 802–805 (1994).
18. Chikashige, Y. et al. *Science* **264**, 270–273 (1994).
19. Bohmann, K., Ferreira, J., Santama, N., Weis, K. & Lamond, A. I. *J. Cell Sci. (suppl.)* **19**, 107–113 (1995).
20. Li, P. & McLeod, M. *Cell* **87**, 869–880 (1996).
21. Sugimoto, A., Iino, Y., Maeda, T., Watanabe, Y. & Yamamoto, M. *Genes Dev.* **5**, 1990–1999 (1991).
22. Basi, G., Schmidt, E. & Maundrell, K. *Gene* **123**, 131–136 (1993).
23. Mochizuki, N. & Yamamoto, M. *Mol. Gen. Genet.* **233**, 17–24 (1992).
24. McLeod, M. & Beach, D. *EMBO J.* **5**, 3665–3671 (1986).
25. Boyle, W. J., Geer, P. V. d. & Hunter, T. *Meth. Enzymol.* **201**, 110–149 (1991).
26. Heim, R., Cubitt, A. B. & Tsien, R. Y. *Nature* **373**, 663–664 (1995).
27. Hiraoka, Y., Swedlow, J. R., Paddy, M. R., Agard, D. A. & Sedat, J. W. *Semin. Cell Biol.* **2**, 153–165 (1991).
28. Shimoda, C. et al. *J. Bacteriol.* **169**, 93–96 (1987).
29. Takeda, T. & Yamamoto, M. *Proc. Natl Acad. Sci. USA* **84**, 3580–3584 (1987).

Acknowledgements. We thank A. Mayeda and M. Yanagida for discussion and advice, P. Fantes for the *cdc22* clone, H. Okayama for the *rep2* clone, Y. Watanabe for GFP DNA, and D. A. Hughes and P. Nurse for help in preparation of the manuscript. This work was supported by Grants-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan.

Correspondence and requests for materials should be addressed to M.Y. (e-mail: myamamoto@ims.u-tokyo.ac.jp).

Crystal structure of the type-I interleukin-1 receptor complexed with interleukin-1 β

Guy P. A. Vigers, Lana J. Anderson, Patricia Caffes & Barbara J. Brandhuber

Amgen Inc., 3200 Walnut Street, Boulder, Colorado 80301, USA

Interleukin-1 (IL-1) is an important mediator of inflammatory disease. The IL-1 family currently consists of two agonists, IL-1 α and IL-1 β , and one antagonist, IL-1ra. Each of these molecules binds to the type I IL-1 receptor (IL1R)¹. The binding of IL-1 α or IL-1 β to IL1R is an early step in IL-1 signal transduction and blocking this interaction may therefore be a useful target for the development of new drugs. Here we report the three-dimensional structure of IL-1 β bound to the extracellular domain of IL1R (s-IL1R) at 2.5 Å resolution. IL-1 β binds to s-IL1R with a 1:1 stoichiometry. The crystal structure shows that s-IL1R consists of three immunoglobulin-like domains which wrap around IL-1 β in a manner distinct from the structures of previously described cytokine–receptor complexes. The two receptor-binding regions

on IL-1 β identified by site-directed mutagenesis^{2,3} both contact the receptor: one binds to the first two domains of the receptor, while the other binds exclusively to the third domain.

Human IL1R was first cloned from T cells⁴ and its sequence predicted a mature receptor consisting of a 552 amino-acid protein containing an extracellular ligand-binding domain of 319 amino acids having three immunoglobulin-like (Ig-like) domains, a single transmembrane segment, and a 213 amino-acid cytoplasmic domain. IL-1 α , IL-1 β and IL1RA (IL-1 receptor antagonist) all bind to IL1R and although they share only about 25% amino-acid sequence identity, these molecules all have the same β -trefoil, 12-stranded β -barrel structure⁵. In order to further our understanding of IL-1–IL1R interactions we have solved the crystal structure of recombinant human IL-1 β bound to recombinant human s-IL1R. The current structure has an *R*-factor of 23.2% at 2.5 Å resolution.

The IL-1 β –s-IL1R complex structure (Fig. 1) was solved using a combination of molecular replacement and multiple isomorphous replacement (MIR) phasing methods (Table 1). The complex has rough dimensions of 97 Å $\times$ 52 Å $\times$ 35 Å with one s-IL1R molecule wrapping around the IL-1 β molecule to contact both the receptor binding sites previously identified by structure: function studies^{2,3}. As predicted by its sequence, s-IL1R has three Ig-like domains (see Fig. 1). All three Ig-like domains contain a conserved pair of cysteines and a tryptophan residue about 14 residues after the first conserved cysteine of each domain. Domain 3 is of the c-type, seven-stranded variety⁶, similar to the immunoglobulin (Ig) constant domain whereas domains 1 and 2 are more unusual topologically. In both of these domains, strand a is strand-switched to the 'front' face of the barrel, as seen in some v-type Igs, but the c' strand is very short (as in constant domains) and the c'' strand is not detectable. Strands d and e are also extremely short, and are comparable in length to those seen in the cell-adhesion molecule *N*-cadherin⁷ (Protein Database (PDB) entry 1NCG). Furthermore, strands c, d and e are on the distal side of domains 1 and 2 from the IL-1 β molecule, and are not involved in IL-1 binding. Thus, all three domains of the IL1RI are most simply classified as c-type. Domains 1 and 2 are closely juxtaposed, with a disulphide bond (C104–

C147) holding them together, whereas domain 3 is connected by a 5-residue linker (L201–K205). Domain 3 provides a 'lid' which covers most of the top of the IL-1 β β -barrel, whereas domains 1 and 2 form a groove which binds to the lower rim of the barrel.

The structure of IL-1 β does not change substantially upon binding. The r.m.s. deviation in position for all non-hydrogen atoms is 1.76 Å between the starting structure of IL-1 β ⁸ and the refined complex structure. Maximum differences are seen at Q32 (r.m.s.d. = 5.6 Å), K88 (r.m.s.d. = 5.5 Å), Q141 (r.m.s.d. = 4.3 Å) and the C-terminal S153 (r.m.s.d. = 9.6 Å). Of these residues, Q32 is involved in receptor binding, whereas the other three are in flexible loops of the molecule. Minimum differences are seen in residues such as A59, L60, F112 and F133 (all about 0.2 Å r.m.s.d.) which are involved in the packing of the hydrophobic core of the IL-1 β ³.

Extensive site-directed mutagenesis work has delineated two receptor-binding sites on IL-1 α and IL-1 β and one on IL1RA. Figure 2 shows that on IL-1 β the two binding sites correspond extremely well with residues buried in the receptor: ligand interface. For example, Gln 15 (shown in blue in the middle of Fig. 2) lost 100% of its IL1R binding activity when mutated to glycine, and 116 Å² of its solvent-accessible surface is buried upon binding s-IL1R. In addition, numerous residues, which showed a loss of 30–50% of their receptor binding activity upon conservative mutation, are located on the edges of the binding sites and show corresponding changes in computed solvent accessibility upon receptor binding. These results demonstrate the utility of mutagenesis studies for mapping ligand: receptor contacts.

The binding site common to all three IL-1 family members, on the 'side' of the IL-1 β β -barrel (site A), makes contacts with β -strands a1, a2, b2 and the b2–c2 loop in domains 1 and 2 of the receptor (Fig. 1) and consists of residues 11, 13–15, 20–22, 27, 29–36, 38, 126–131, 147 and 149. The critical binding residues Arg 11 and Gln 15 contact domain 2 whereas His 30 and Gln 32 bind in the domain 1–2 junction. The surface of the receptor at this binding site has a number of pockets that will accept side chains of IL-1 β . Thus a significant component of the binding energy at this site probably

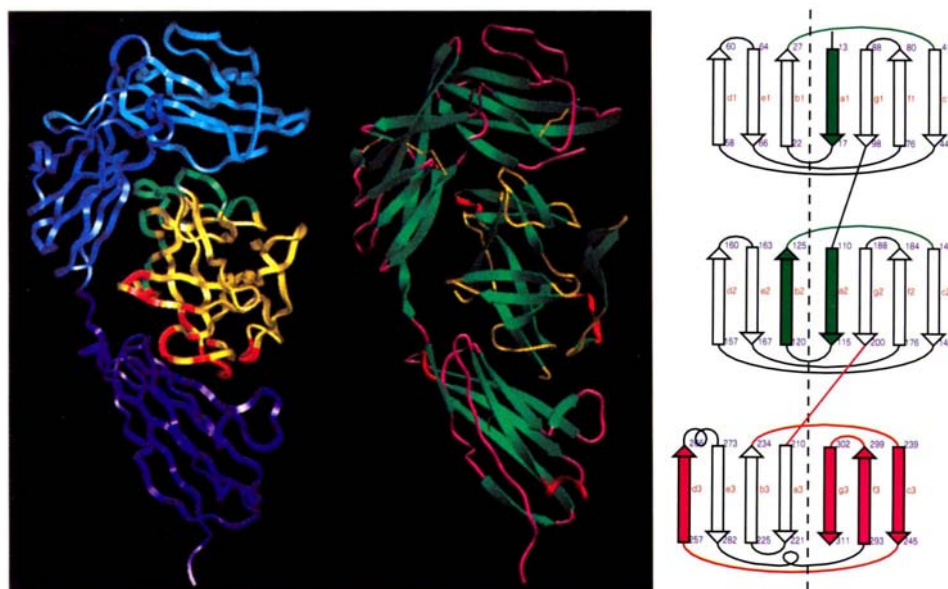


Figure 1 Left, Ribbon diagram of s-IL1R complexed to IL-1 β . Domains 1, 2 and 3 of s-IL1R are coloured light, medium and dark blue, respectively. IL-1 β is yellow, with site A residues in green and site B residues in red. Middle, Ribbon diagram of the structure of IL-1 β bound to s-IL1R. The β -strands are shown as arrowed ribbons in green, α -helices are red, and the connecting loops are purple. The structure is

oriented so that the carboxy terminus of s-IL1R and the cell membrane, if present, are at the bottom of the picture. Right, Topology diagram of s-IL1R. The IL-1 β -binding elements in site A and site B are coloured green and red, respectively. Secondary structural elements are not to scale.

comes from van der Waals contacts between IL-1 β and s-IL1R. Interestingly, the position of Gln 32 has moved by more than 5 Å from its position in the unliganded X-ray structure, and now fits in a deep pocket on the s-IL1R (Fig. 3).

Site B, the second binding site identified on IL-1 α and IL-1 β but not IL1RA, is on the top of the β -barrel and is formed by residues 1–4, 6, 46, 48, 51, 53–54, 56, 92–94, 103, 105–106, 108, 109, 150 and 152. Site B makes contacts only with domain 3 of s-IL1R. Previous work from our group⁵ suggested that site B of IL-1 β was composed of a 'horseshoe' of hydrophilic residues surrounding several hydrophobic residues. As shown in Fig. 3, both the hydrophilic residues (Arg 4, Gln 48, Glu 51, Asn 53, Lys 93, Glu 105 and Asn 108) and the hydrophobic residues (Leu 6, Phe 46, Ile 56 and Phe 150) do indeed contact IL1R over a large and relatively flat surface (formed by β -strands c3, d3, f3, g3 and the b3–c3, c3–d3, f3–g3 and g2–a3 loops), which is complementary to IL-1 β in hydrophobicity. Thus, the binding energy of site B appears to be more dependent upon hydrophobic and hydrophilic interactions than in site A. Because IL1RA lacks the distinctive 'horseshoe' feature of IL-1 β , formation

of the IL1RI-IL1RA complex will presumably not make the same energetically favourable site B contacts seen in this structure.

Calculating the changes in solvent-accessible area of the IL-1 β on binding s-IL1R, 1,087 Å² are buried in site A over 25 residues and 1,001 Å² in site B over 21 residues, for a total of 2,088 Å² overall. In addition to the hydrophilic–hydrophobic interactions described above (which are particularly apparent for site B), both binding sites also make extensive use of salt bridges and hydrogen bonds. Site A and site B have 10 and 13 salt bridges, respectively, and both sites also contain seven intramolecular hydrogen bonds (as deduced from the X-ray structure). It is interesting to note that in this complex, the ligand-binding interactions are with the β -strands in the faces of the Ig-like domains, not with the 'elbow' formed by the loops between the domains, as seen in the structures of interferon- γ –interferon- γ receptor⁹, human growth hormone–human growth hormone-binding protein¹⁰ (PDB entry 3HHR) and the erythropoietin mimetic peptide–erythropoietin-binding protein¹¹ (PDB entry 1EBP). Also, in previously determined cytokine–receptor structures, whereas the cytokine might exist as a monomer

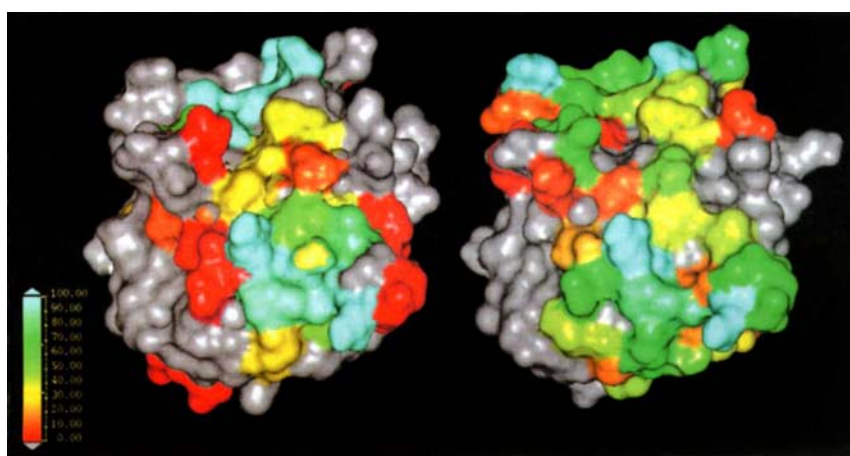


Figure 2 Surface representations of IL-1 β . Left, Structure of unbound IL-1 β ⁸; spectral colouring based on site-directed mutagenesis experiments. Blue residues lost 100% of IL1R binding activity, red residues lost no activity, and grey residues were not done. Right, Structure of IL-1 β from the complex coloured to

show the change in solvent-accessible area upon complex formation¹⁸. Blue residues bury 100 Å² or greater, and red residues bury 2 Å². Grey residues bury less than 2 Å². Site B is at the top and site A faces the viewer.

Table 1 MIR data

Data set	Resolution (Å)	R_{sym} (last shell)	Completeness (last shell)	$\langle L/s \rangle$ (last shell)	Sites (n)	R_{cullis_a}	R_{cullis_c}	$R_{\text{cullis}_{\text{ano}}}$	Phasing power _a	Phasing power _c
Native*	2.5	6.6 (10.4)	96.2 (94.9)	28.03 (10.94)	–	–	–	–	–	–
HgCl ₂ _1	3.0	7.5 (27.2)	79.3 (29.5)	15.11 (5.46)	2	50.0	48.0	70.0	2.0	1.7
HgCl ₂ _2	2.8	8.1 (31.4)	97.4 (98.5)	13.74 (4.51)	2	64.0	62.0	70.0	1.5	1.2
PtCl ₁	3.0	6.1 (27.6)	84.4 (91.6)	15.31 (4.55)	5	76.0	66.0	60.0	1.1	1.0
PtCl ₂	3.0	7.8 (30.7)	76.5 (80.4)	15.07 (5.33)	5	75.0	68.0	80.0	1.2	1.1
C71S_1†	3.0	7.2 (19.8)	74.7 (80.3)	14.15 (6.23)	1	63.0	62.0	80.0	1.6	1.2
C71S_2†	2.5	6.6 (19.6)	93.0 (95.2)	16.66 (5.79)	1	72.0	71.0	80.0	1.4	1.1
Q126C_1‡	3.0	7.3 (23.7)	78.7 (85.8)	15.93 (5.95)	2	74.0	63.0	80.0	1.1	1.0
Q126C_2‡	2.5	4.9 (18.3)	94.3 (97.2)	17.90 (6.85)	2	86.0	78.0	80.0	0.9	0.8

* Data merged from two data sets.

† IL-1 β mutated so that cysteine 71 is serine.

‡ IL-1 β mutated so that cysteine 71 is serine and glutamine 126 is cysteine.

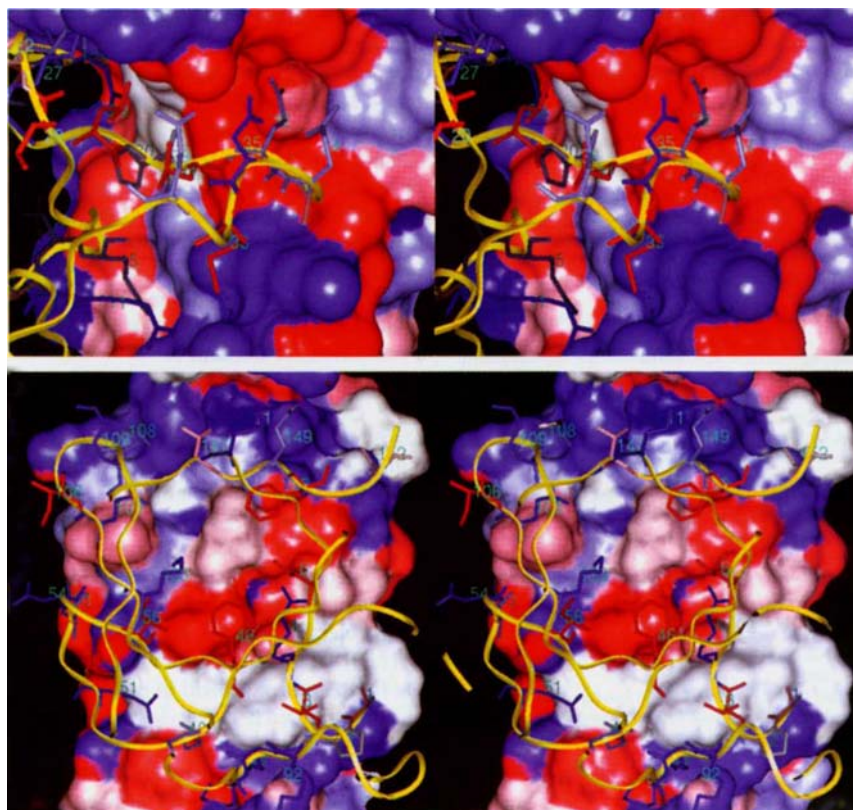


Figure 3 Stereoscopic surface representations of s-IL1R and ribbon diagrams of IL-1 β . Top, Site A. Bottom, Site B. Side chains in IL-1 β and the surface of s-IL1R are

coloured red for hydrophobic residues and purple for hydrophilic residues. The IL-1 β backbone ribbon is yellow.

(hGH), dimer (IFN γ) or trimer (TNF β)¹², in every case two or more receptors bound to the cytokine and each receptor made a single contiguous surface contact with the cytokine. The s-IL1R binding motif thus provides a new model for receptor–cytokine interactions.

IL-1 signalling is believed to result from the formation of a ternary complex consisting of an IL-1 agonist, IL1R and IL1R accessory protein (IL1RAcP)¹³. When IL1RA binds IL1R, the ternary complex is not formed and signalling does not occur. Formation of the ternary complex therefore presumably requires an epitope that is created when IL1RI binds IL-1 β but not IL1RA. Whether the epitope is created by formation of the receptor–ligand interface, or by conformational changes of IL1RI on complex formation, has yet to be determined. Site-directed mutagenesis of additional surface residues in IL-1 and IL1R, and solving the structure of the ternary complex, will further advance our understanding of the IL-1 family members' activities. □

Methods

Purification and crystallization. Recombinant human s-IL1R (residues 1–315, GenBank numbering) was expressed in *Escherichia coli*, refolded and purified using an IL1RA affinity column, MonoQ ion-exchange column and gel filtration column (B.J.B. *et al.*, manuscript in preparation). Wild-type and mutants of IL-1 β used for heavy-atom phasing were generated and purified as described previously³. The mercury chloride derivative was found to derivatize the cysteines only in IL-1 β , therefore cysteine mutants of IL-1 β were made to augment the phase information available. All mutants that were crystallized and derivatized were isomorphous with native, wild-type complex crystals. The s-IL1R was incubated overnight in a fivefold molar excess of IL-1 β , purified over a Superdex 75 column and concentrated to 5 mg ml⁻¹. Crystals were grown by hanging-drop diffusion against 1.8 M ammonium sulphate, 100 mM MES pH 6.0 at 0–4 °C. The crystals grew to about 0.4 × 0.4 × 3 mm in 1–2 weeks.

Data collection and phasing. Initial characterization of the crystals and searches for heavy-atom derivatives were performed on room-temperature crystals in an artificial mother liquor containing 2.8 M ammonium sulphate, 100 mM MES pH 6.0. However, the crystals proved to be very radiation sensitive at room temperature and all data sets for MIR phasing and refinement were therefore collected on cryo-cooled crystals. Crystals were cryoprotected with 25% glucose, 1.3 M ammonium sulphate, 100 mM MES pH 6.0 and diffracted to 2.5 Å resolution. The crystals were of space group C2 with unit cell dimensions $a = 146.95$ Å, $b = 68.45$ Å, $c = 65.87$ Å, $\alpha = \gamma = 90.0^\circ$, $\beta = 108.95^\circ$. X-ray data were collected on an R-axisII area detector, reduced using Denzo and Scalepack¹⁴ and processed using the CCP4 suite of programs including MLPHARE and DM¹⁵. The position of the IL-1 β molecule in the complex was determined by molecular replacement with X-plor 3.1 from the PDB coordinates pdb2ilb.ent⁸.

Structure refinement. The s-IL1R structure was traced and rebuilt using program O¹⁶. Refinement by conjugate gradient and simulated annealing was performed in X-PLOR 3.1¹⁷ for all reflections with $F > 2\sigma(F)$ from 15.0 to 2.5 Å ($n = 20,723$). All solvent-accessibility calculations were performed in X-PLOR using the method of Lee and Richards¹⁸, using a 1.4 Å radius probe and assuming that the ligand and receptor structures do not change significantly upon binding. Potential hydrogen bonds were determined using the 'Hbonds_all' facility in O. Salt bridges were assigned in X-PLOR by picking all intermolecular oxygen–nitrogen pairs separated by 3.4 Å or less. Secondary structure assignments were made with the YASSPA algorithm in O and pictures were generated using Setor¹⁹. The current structure contains 31 ordered water molecules and a bulk solvent correction (sol density = 0.40, sol rad = 0.25) was used for the glucose cryoprotectant. The R -factor of the current structure is $R = 23.2\%$ with a free R -factor of 33.1%. In the MIR maps, part of the amino terminus of the first Ig-like domain showed only weak density. For this reason residues 1–8 and 30–41 are not modelled, and the 40–54 loop is poorly defined. Fortunately, this region is distal to the IL-1 binding site, and so does not significantly affect the conclusions of the study. The r.m.s. deviations from ideality are 0.012 Å for bond lengths and 1.98° for bond angles. Running

Procheck on the refined structure showed that four residues are in disallowed regions of the Ramachandran plot. Two of these residues (E11 and Q53) are in the poorly defined region of the molecule whereas the other two (L186 and N246) are at the apex of very tight loops.

Received 18 November 1996; accepted 20 January 1997.

1. Dinarello, C. A. Biologic basis for interleukin-1 in disease. *Blood* **87**, 2095–2147 (1996).
2. Labriola-Thompkins, E. *et al.* Identification of the discontinuous binding site in human interleukin 1 β for the type I interleukin 1 receptor. *Proc. Natl Acad. Sci. USA* **88**, 11182–11186 (1991).
3. Evans, R. J. *et al.* Mapping receptor binding sites in interleukin (IL)-1 receptor antagonist and IL-1 β by site-directed mutagenesis. *J. Biol. Chem.* **270**, 11477–11483 (1995).
4. Sims, J. E. *et al.* Cloning the interleukin 1 receptor from human T cells. *Proc. Natl Acad. Sci. USA* **86**, 8946–8950 (1989).
5. Vigers, G. P. A. *et al.* X-ray structure of interleukin-1 receptor antagonist at 2.0 Å resolution. *J. Biol. Chem.* **269**, 12874–12879 (1994).
6. Bork, P., Holm, L. & Sander, C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320 (1994).
7. Shapiro, L. *et al.* Structural basis of cell-cell adhesion by cadherins. *Nature* **374**, 327–337 (1995).
8. Priestle, J. P., Schaer, H. P. & Grutter, M. G. Crystallographic refinement of interleukin 1 beta at 2.0 Å resolution. *Proc. Natl Acad. Sci. USA* **86**, 9667–9671 (1989).
9. Walter, M. R. *et al.* Crystal structure of a complex between interferon- γ and its soluble high-affinity receptor. *Nature* **376**, 230–235 (1995).
10. de Vos, A. M., Ultsch, M. & Kossiakoff, A. A. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306–312 (1992).
11. Livnah, O. *et al.* Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 Å. *Science* **273**, 464–471 (1996).
12. Banner, D. W. *et al.* Crystal structure of the soluble human 55 kd TNF receptor-human TNF beta complex: implications for TNF receptor activation. *Cell* **73**, 431–445 (1993).
13. Greenfeder, S. A. *et al.* Molecular cloning and characterization of a second subunit of the interleukin 1 receptor complex. *J. Biol. Chem.* **270**, 1357–13765 (1995).
14. Otwinowski, Z. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) (SERC Daresbury Laboratory, Warrington, UK, 1993).
15. CCP4: Collaborative Computational Project No 4, Daresbury UK. *Acta Crystallogr. D* **50**, 760 (1994).
16. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
17. Brunger, A. T., Kuriyan, J. & Karplus, M. Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458–460 (1987).
18. Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400 (1971).
19. Evans, S. V. SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134–138 (1993).

Acknowledgements. We thank T. Osslund, R. Syed, V. Carperos and C. Kundrot for use of data collection equipment. We thank J. Rizzi and T. Kohno for stimulating discussions.

Correspondence and requests for materials should be sent to B.J.B. (e-mail: barbara.brandhuber@amgen.com). Atomic coordinates have been deposited with the Bookhaven Protein Data Bank, PDB ID code 1ltb.

A new cytokine-receptor binding mode revealed by the crystal structure of the IL-1 receptor with an antagonist

Herman Schreuder*†, Chantal Tardif*†, Susanne Trump-Kallmeyer*, Adolfo Soffientini‡, Edoardo Sarubbi‡, Ann Akeson§, Terry Bowlin§, Stephen Yanofsky|| & Ronald W. Barrett||

* Marion Merrell Dow Research Institute, 67080 Strasbourg Cedex, France

† Lepetit Research Centre, 21040 Gerezano, Italy

§ Hoechst Marion Roussel Inc., Cincinnati, Ohio 45215-6300, USA

|| Affymax, Palo Alto, California 94304, USA

Inflammation, regardless of whether it is provoked by infection or by tissue damage, starts with the activation of macrophages which initiate a cascade of inflammatory responses by producing the cytokines interleukin-1 (IL-1) and tumour necrosis factor- α (ref. 1). Three naturally occurring ligands for the IL-1 receptor (IL1R) exist: the agonists IL-1 α and IL-1 β and the IL-1-receptor antago-

nist IL1RA (ref. 2). IL-1 is the only cytokine for which a naturally occurring antagonist is known. Here we describe the crystal structure at 2.7 Å resolution of the soluble extracellular part of type-I IL1R complexed with IL1RA. The receptor consists of three immunoglobulin-like domains. Domains 1 and 2 are tightly linked, but domain three is completely separate and connected by a flexible linker. Residues of all three domains contact the antagonist and include the five critical IL1RA residues which were identified by site-directed mutagenesis³. A region that is important for biological function in IL-1 β , the 'receptor trigger site', is not in direct contact with the receptor in the IL1RA complex. Modelling studies suggest that this IL-1 β trigger site might induce a movement of domain 3.

A soluble form of the IL-1 receptor, comprising residues 1–311, was expressed in insect cells and purified by affinity chromatography⁴. The soluble receptor was mixed with recombinant IL1RA at a ratio of 1:1 and was crystallized from 30% PEG3350 at pH 7.0–7.5⁴. The structure has been solved by a combination of molecular replacement and heavy-atom derivatives and has been refined at 2.7 Å resolution to an *R* factor of 21.1% (*R*_{free} = 31.4%) with excellent geometry (see Methods and Table 1). A representative sample of the original squashed multiple isomorphous replacement (MIR) electron density map is given in Fig. 1.

The IL1R molecule resembles a question mark which curls round the IL1RA molecule held in its centre (Fig. 2a). The IL1RA consists of a six-stranded β -barrel closed at one side by three β -hairpin loops and is similar to the structure of the free molecule we solved previously⁵. The root-mean-square (r.m.s.) differences in alpha carbon (α) positions between bound and free IL1RA are 0.7 Å and 0.8 Å, respectively for the two independent IL1RA molecules present in the crystals of the free IL1RA. The largest differences in the antagonist occur in the loop comprising residues 51–54. This loop interacts with the third IL1R domain but is partially disordered in the free antagonist. Upon binding to IL1R, there is a rotation of the side chain of Glu 52 away from the receptor, a rotation of the side chain of His 54 towards the receptor, and a change in conformation from *cis* to *trans* for Pro 53. Other differences in atomic position greater than 5.0 Å involve solvent-exposed, flexible side chains, namely those of Lys 64, Glu 75, Arg 77, Lys 93, Lys 96, Asp 138, Glu 139 and Lys 145.

As predicted⁶, the folding of the three IL1R domains, consisting of a sandwich of two β -sheets, resembles the folding of immunoglobulin domains. The topology of domains 1 and 2 (residues 1–94 and 106–198) is distantly related to the telokin I-set⁷ topology, although it diverges significantly from the canonical I-set structures. A β -bulge is present between residues 8 and 11 in strand a of domain 1. The bed sheets in domains 1 and 2 are very short, with strand lengths of only 3–4 residues. Strands g and f in domain 2 are long and have extensive contacts with domain 1. Residues involved are Glu 18, Gln 50, His 55 and Val 64 of domain 1, and Tyr 182, Leu 183, Tyr 187, Pro 188 and Thr 190 of domain 2. Domains 1 and 2 are linked through a well-defined β -turn between Pro 98 and Asn 99 which lies on top of the agfc' sheet (Fig. 2b). Domains 2 and 3 are connected through an extended, flexible linker.

The electron density maps clearly indicated four carbohydrate-binding sites at Asn 173, 213, 229 and 243. The electron density quickly disappears after the first *N*-acetylglucosamine unit, indicating that most sugar units are disordered. The first attachment site, Asn 173, is at the surface of the second domain and its sugar chain is facing the solvent. The other three attachment sites (Asn 213, 229 and 243) are at the surface of domain 3, and in the natural situation their sugar chains will be close to the cell membrane. Interactions of these sugar chains with the membrane could help to stabilize the orientation of the IL1R with respect to the membrane.

Alignment of the sequence of the extracellular part of the type-I IL1R with homologous proteins (Fig. 3) reveals a number of conserved sequence motifs: the conserved disulphide link and

† Present addresses: Hoechst AG, Building G865A, 65926 Frankfurt, Germany (H.S.); Synthelabo Biomoléculaire, 67080 Strasbourg Cedex, France (C.T.); Parke Davis Pharmaceutical Research, Ann Arbor, MI 48105, USA (S.T.-K.); Biochem Therapeutics, Laval, Quebec H7V4A7, Canada (T.B.).

tryptophan, present in the core of many immunoglobulin domains, are absolutely conserved with the exception of IL1R-related protein (IL1RR). A motif that appears to be specific for the IL1R family domains 1 and 2 is a proline immediately after the first cysteine of the conserved core disulphide link. This proline disrupts strand b and is responsible for the very short β sheets observed in domains 1 and 2 of the IL1R. The presence of Asp 20 and Arg 22 on strand b1 at positions usually occupied by highly conserved hydrophobic residues is unusual. The two residues form, together with Ser 14, a hydrophilic cluster in what would otherwise be the hydrophobic core. Almost perfectly conserved is a D_xG_xY_xC motif at the beginning of strands f1 and f2. This motif indicates the presence of a D4 tyrosine corner and is found in many immunoglobulin variable domains⁸. The cysteine of the motif (76 and 176 in IL1R) is part of the core disulphide link. The hydroxyl of the tyrosine makes

a hydrogen bond with the carbonyl oxygen of residue $n - 4$, usually Asp. In domain 1, this Asp makes a salt link with a conserved Arg at the beginning of strand d. The conserved Gly residue in the motif allows the main chain to curve around the tyrosine side chain. Although a D4 tyrosine corner is not present in IL1R domain 3 (the Tyr has been replaced by a Phe and there is a one-residue insertion), the overall conformation of the main chain is similar. Glu 18, at the beginning of strand b1, and Thr 190, halfway along strand g2, make a hydrogen bond in IL1R. These residues are conserved in all sequences, except FGR4, suggesting that the hydrogen bond between them is conserved as well. This would mean that the orientation of domain 1 with respect to domain 2 is similar in these proteins. We conclude from the presence of these common sequence motifs that the proteins listed in Fig. 3 are all similarly folded.

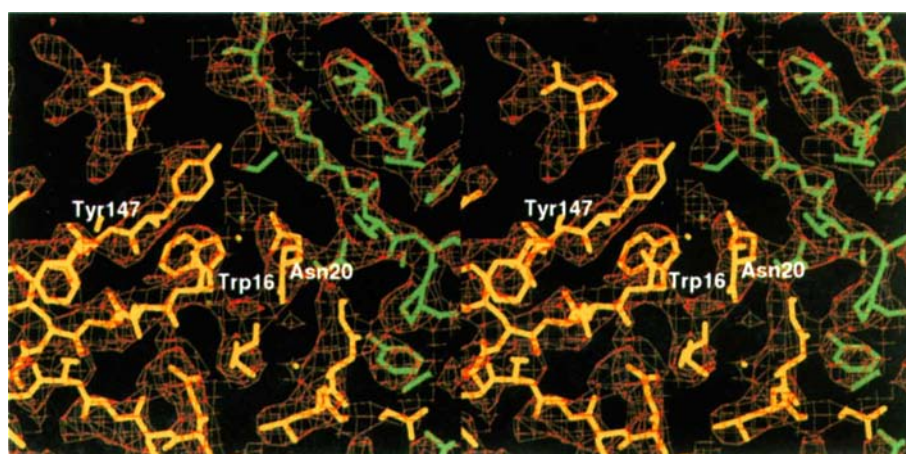


Figure 1 Stereo view of the squashed MIR map contoured at 1σ . Shown is the receptor-antagonist interface in the region around Trp 16 of the antagonist. The refined model of the IL1RA is shown in yellow; IL1R is in green.

Table 1 Crystallographic data

Data collection and phasing statistics				
Data set	Native	Hg(CN) ₂	K ₂ PtCl ₄	Hg(CN) ₂ + K ₂ PtCl ₄
Measured reflections	73,407	71,530	66,081	43,056
Unique reflections	15,631	15,725	14,086	10,835
Resolution (Å)	2.7	2.7	2.8	3.0
Completeness (%)	97.2	98.2	98.5	92.4
$R_{\text{merge}}(\%)^*$	6.0	6.0	6.8	7.5
$R_{\text{iso}}(\%)^\dagger$		22.3	14.8	39.7
Binding sites		1	2	3
Phasing power‡		1.29	0.91	0.68
$R_{\text{cullis}}^\S$		0.58	0.63	0.69
Mean figure of merit	0.47			
Refinement statistics				
Number of protein atoms		3,638		
Number of sugar atoms (NAG)		56		
Number of water molecules added		86		
Resolution range		8.0–2.7 Å		
Data cutoff		None		
R factor (number of reflections)		21.1% (13,487)		
R_{free} (number of reflections)		31.4% (1,525)		
R_{total} (number of reflections)		22.2% (15,012)		
Average temperature factor		45.2 Å ²		
R.m.s. deviations with respect to ideal (target) values				
Bond lengths		0.004 Å		
Bond angles		1.2°		
Dihedrals		26.5°		
Impropers		1.0°		

NAG, *N*-acetylglucosamine.

* $R_{\text{merge}} = \sum_{hkl} |I_{hkl} - \langle I_{hkl} \rangle| / \sum_{hkl} I_{hkl}$.

† $R_{\text{iso}} = \sum_{hkl} |F_{\text{obs}} - F_{\text{calc}}| / \sum_{hkl} |F_{\text{obs}}|$.

‡ Phasing power = $\sum_{hkl} |F_{\text{obs}}| / \sum_{hkl} |F_{\text{calc}}|$ lack of closure.

§ $R_{\text{cullis}} = (\text{lack of closure}) / (\text{isomorphous differences})$ for centric reflections, as calculated by HEAVY²³.

A large area of $1,774 \text{ \AA}^2$ of the IL1RA surface and $1,767 \text{ \AA}^2$ of the receptor surface (calculated by using DSSP⁹) gets buried upon receptor binding (Fig. 4). This area corresponds to about 20% of the total surface of IL1RA and may explain the tight binding ($\sim 10^{-9} \text{ M}$)⁶ of the antagonist to the receptor.

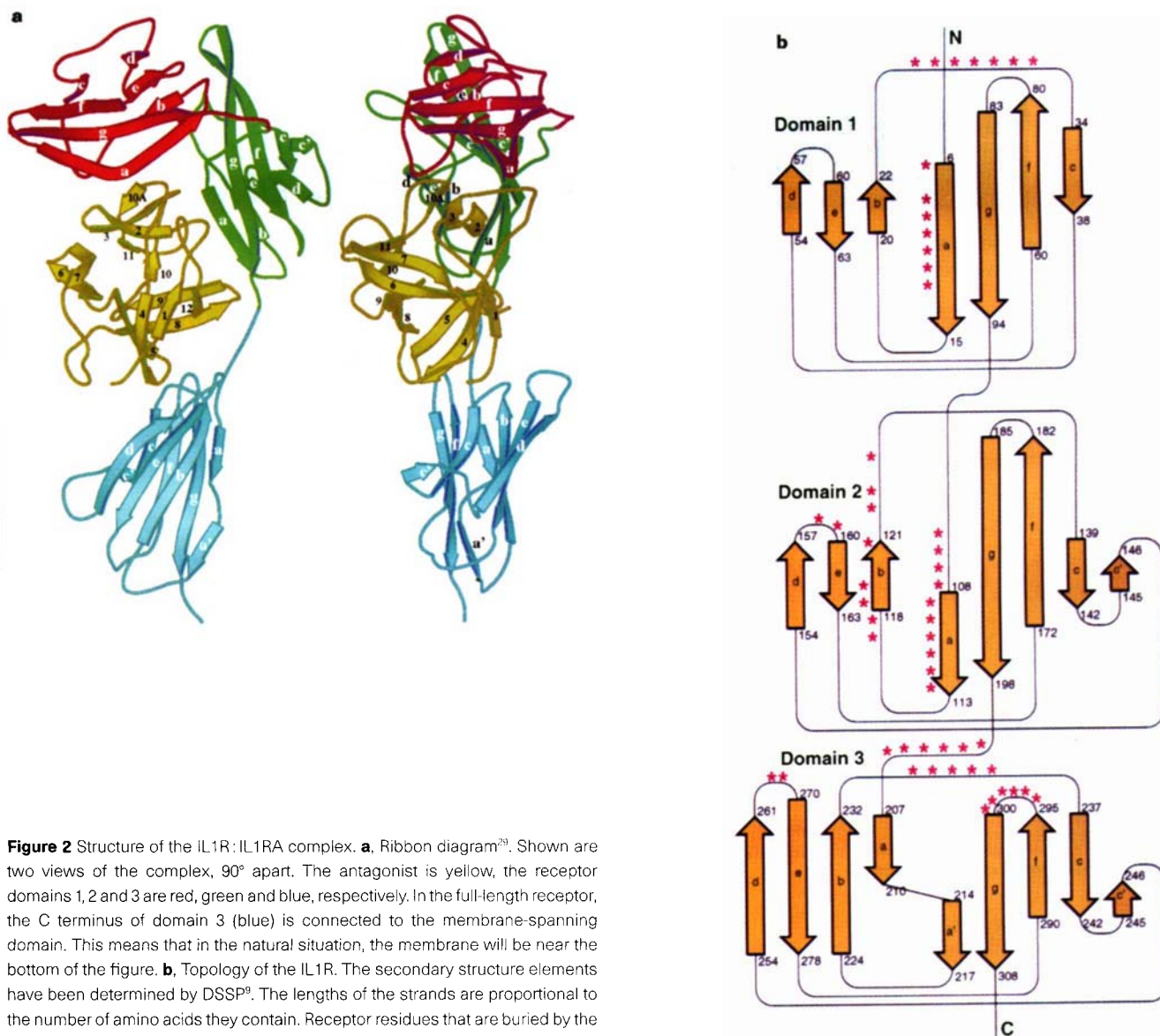
The IL1RA molecule interacts with all three receptor domains. Leu 25, Arg 26, Asn 27, Glu 29, Leu 42 and Pro 130 contact strand a and loop bc in the N-terminal half of receptor domain 1. The loop comprising residues 34–39, which is fully exposed in the crystal structure of free IL1RA⁵, fits in the cleft between domains 1 and 2. Arg 14, Trp 16, Val 18, Asn 19, Gln 20, Ala 127, Asp 128, Tyr 147 and Gln 149 contact strands of both sheets of the β -sandwich of domain 2. Finally, Ser 8, Lys 9, Pro 50, Ile 51, Pro 53, His 54, Leu 56, Glu 150 and Asp 151 contact loops bc and fg at the top of domain 3 (Fig. 2).

The interactions between the IL1R and the antagonist are mainly polar in character. We observe only three hydrophobic contacts (Table 2). Surprisingly, despite the many charged residues on the IL1R and the antagonist, we find only two direct and one indirect

Table 2 Selected interactions ($d \leq 3.5 \text{ \AA}$) between the IL1R and IL1RA

Side-chain–side-chain interactions		Interactions involving the main chain	
Receptor	Antagonist	Receptor	Antagonist
Salt links		Receptor side chain to antagonist main chain	
Glu 8	Arg 26	Lys 9	Leu 25
Arg 268	Glu 150	Glu 126	Ala 127
Hydrogen bonds		Tyr 124	Tyr 34
Tyr 124 OH	Asp 128 o δ 2	Arg 160	Val 18
Ser 235 O γ	Gln 11 O ϵ 1	Receptor main chain antagonist side chain	
Hydrophobic contacts		Ile 11	Asn 39*
Ile 10	Leu 42	Val 13	Gln 36*
Leu 198	Trp 16	Ala 106	Gln 36
Pro 113	Tyr 147	Lys 111	Trp 16
		Lys 111	Gln 20*
		Gly 119	Gln 20
		Receptor main chain to antagonist main chain	
		Ile 107	Gly 37
		Asn 296	Pro 53
		Thr 297	His 54

* These residues make two hydrogen bonds: one from the side-chain amide to the main-chain carbonyl, and one from the side-chain carbonyl to the main-chain amide.



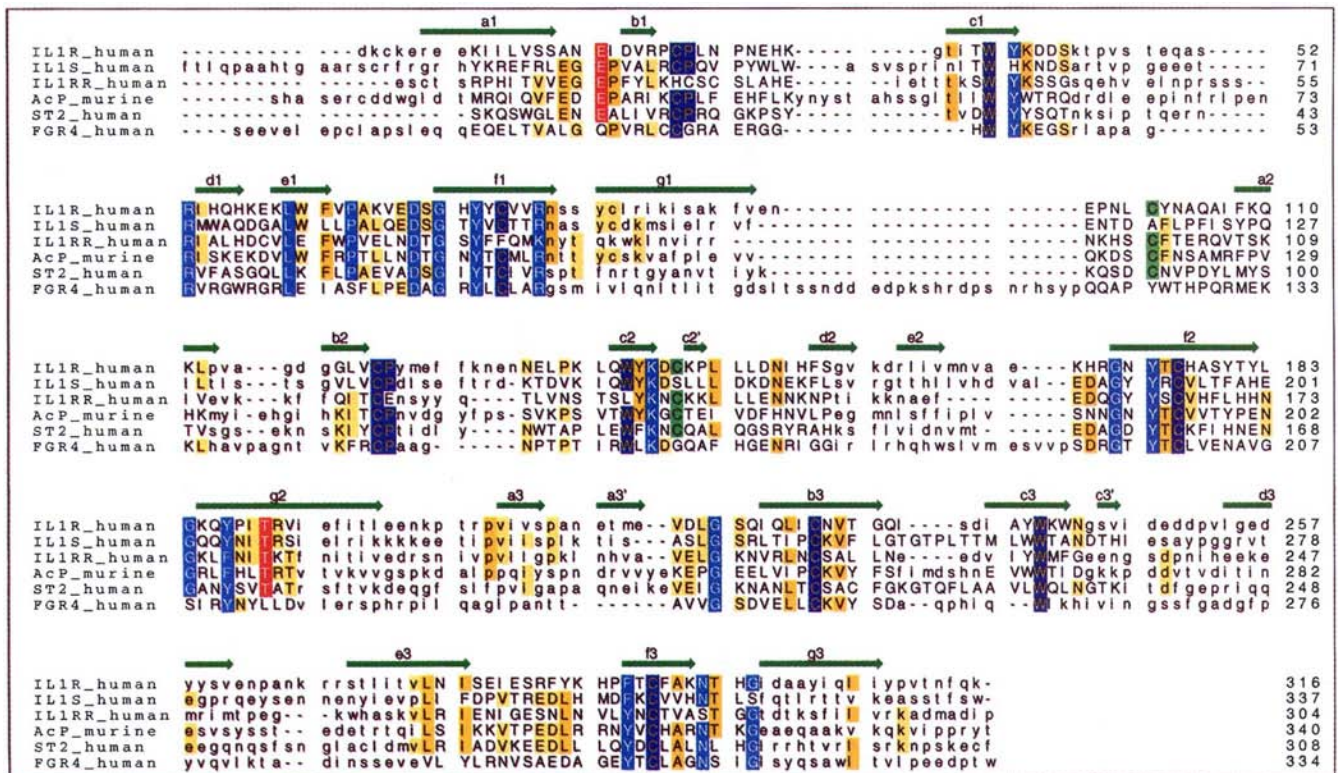


Figure 3 Sequence alignment of the extracellular portion of the IL1R with the extracellular portions of 5 homologous proteins. IL1R, type-I IL-1 receptor; IL1S, type-II IL-1 receptor; IL1RR, IL-1-receptor-related protein; AcP, IL-1-receptor accessory protein; ST2, ST2 protein; FGR4, fibroblast growth factor receptor 4. The cysteines and tryptophans that are conserved in most immunoglobulin folds are indicated in dark blue. Also highlighted are the prolines immediately following

conserved cysteines, which appear to be conserved in the particular immunoglobulin fold of IL-1 domains 1 and 2. Highlighted in green are 5 conserved cysteine pairs which, based on the IL1R structure, are likely to form disulphide bonds. Other highly conserved residues are indicated in blue. Conserved glutamic acids and threonines, which are probably involved in a hydrogen bond between domains 1 and 2, are indicated in red.

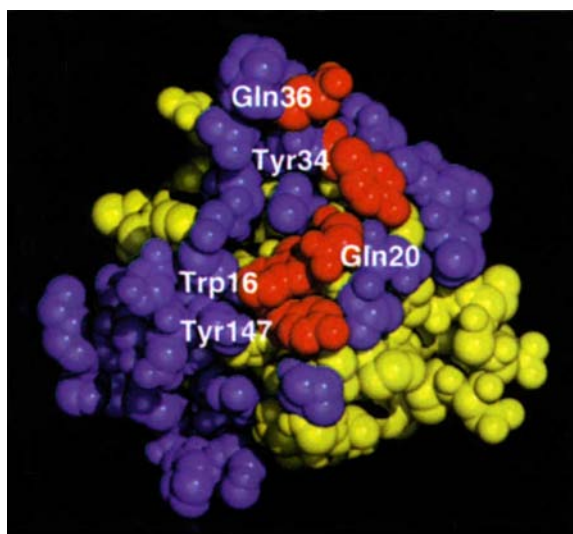


Figure 4 Space-filling model of bound IL1RA. Residues shown by mutagenesis to be important for receptor interaction³ are in red, other residues that are buried by the receptor are in blue, and those that do not interact with the receptor are in yellow. The five residues defined by site-directed mutagenesis are right in the centre of the area that is buried by the receptor in the crystal structure. However, the contact area defined by the mutagenesis studies is only 29% of the area defined by the crystal structure, suggesting that many of the contacting residues contribute less significantly to the overall binding energy.

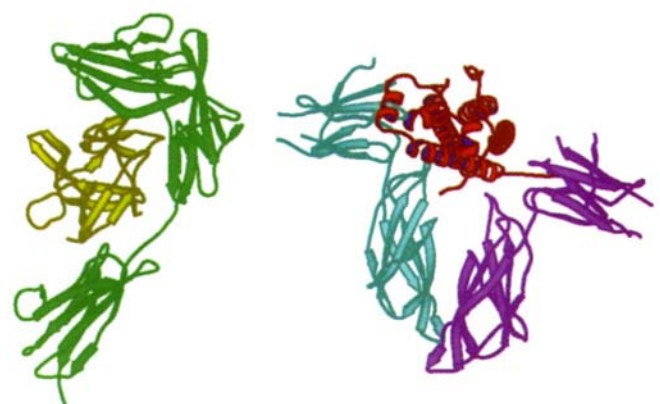


Figure 5 Comparison of the IL1R (left) and HGHR (right) complexes. IL1RA is shown in yellow, IL1R in green, HGHR in red and the two HGHR molecules in blue and magenta. The two receptors use opposite faces of the immunoglobulin domains to bind the ligands.

salt link (Lys 11R and Glu 43A are linked through a water molecule). Some oppositely charged side chains (Lys 129R–Glu 126A, Lys 295R–Glu 52A, Lys 295R–Glu 90A) are less than about 8 Å apart, but do not seem to make any direct interactions. It may be that for these residue pairs, the gain in electrostatic energy is offset by a loss of entropy. This idea is supported by a mutagenesis study that shows that one of the salt links in the crystal structure of human growth hormone complexed to its receptor¹⁰ contributes little, if anything, to the overall binding energy, an effect the authors ascribe to entropy loss.

We observe only two side-chain–side-chain hydrogen bonds between the receptor and the antagonist. Most of the hydrogen bonds between antagonist and receptor involve main-chain carbonyl or nitrogen atoms (Table 2). Of special interest are the double hydrogen bonds made by Gln 20, Gln 36 and Asn 39 of the antagonist with the backbone of the receptor, which suggests that these residues may contribute significantly to the overall binding. The potential advantages of using main-chain atoms for receptor–ligand interactions are twofold: the main chain is usually more rigid than the side chains so that the entropy loss for the main chain will be less than for side chains; and interactions involving main-chain atoms depend less on the sequence and more on the overall folding than interactions between side chains, which may explain how the IL1R is able to bind its three natural ligands, IL-1 α , IL-1 β and IL1RA with high affinity, despite the low sequence identity (20–25%) between them.

The crystal structure of the receptor complex fully explains the published mutagenesis studies. IL1RA residues Trp 16, Gln 20, Tyr 34, Gln 36 and Tyr 147, identified as critical for antagonist binding³, are all in direct contact with the receptor (Table 2; Fig. 4). The same residues (Tyr, Trp, Gln) are present in the YWQPWA consensus sequence, identified by affinity screening using phage

display peptide libraries¹¹. The indole side chain of Trp 16 is completely buried by the receptor and makes a hydrogen bond with the carbonyl oxygen of Lys 111. The side chain of Gln 20 makes two hydrogen bonds with the receptor backbone. The side chain of Tyr 34 makes no direct contacts of less than 3.5 Å with the receptor. It does face a hydrophobic region on the receptor surface, consisting of the side chains of Phe 108, Val 121, Pro 123 and Tyr 124. The side chain of Tyr 34 is tightly fixed in position by a hydrogen bond with the carbonyl oxygen of Asn 19 and its carbonyl oxygen makes a hydrogen bond with the hydroxyl group of Tyr 124. The Gln 36 side chain is found in a pocket between receptor domains 1 and 2 and makes three hydrogen bonds with the receptor backbone. These hydrogen bonds and the restricted pocket clearly explain the large loss of binding affinity associated with the Gln 36 $\rightarrow$ Phe mutation. Finally, the hydroxyl of the critical Tyr 147 binds to a network of two fixed water molecules on top of loop bc of receptor domain 2.

The binding of IL1RA in the IL1R complex is fundamentally different from the binding of ligands in the cytokine–receptor complexes whose three-dimensional structures are known. In the tumour necrosis factor- β (TNF- β) complex¹², three elongated receptor molecules are bound to the ligand and the receptor domains do not possess an immunoglobulin fold. In the receptor complexes with human growth hormone (HGHR¹³), interferon- γ (IFN γ R¹⁴) and erythropoietin peptide (EPOR¹⁵), the ligand binds to the outside of the elbow between two immunoglobulin-like domains of the fibronectin type and is sandwiched between two receptor molecules (Fig. 5). In the IL1RA complex, the receptor has three immunoglobulin-like domains and the ligand binds to the inside of the elbow formed by domains 1 and 2.

Another way in which the IL1R differs from HGHR, IFN γ R and many other proteins with immunoglobulin-like domains is in the interactions between its domains. Most immunoglobulin-like

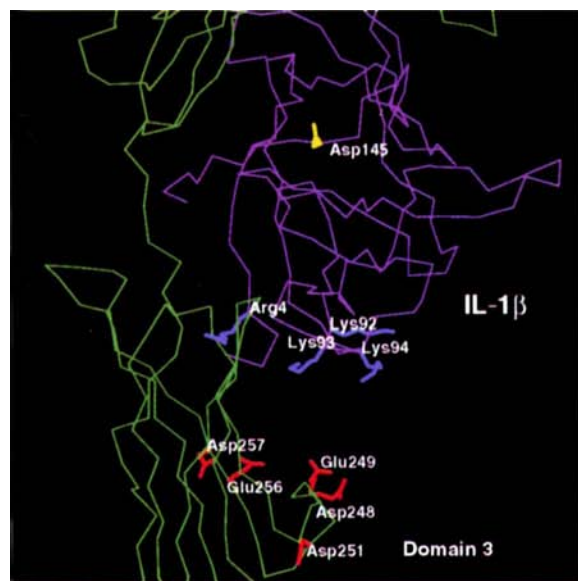


Figure 6 Model obtained after superimposing IL-1 β onto IL1RA in the complex. IL-1 β is indicated in magenta and the receptor in green. After this superposition, a region of positively charged residues (Arg 4, Lys 92, Lys 93 and Lys 94) in IL-1 β is at a distance of $\sim$ 15 Å from a region of negatively charged residues (Asp 248, Glu 249, Asp 251, Glu 256 and Asp 257) in domain 3 of IL1R but do not directly interact with them. Modelling studies (not shown) suggest that upon movement of IL1R domain 3, these two regions interact directly. A definitive answer must come from the crystal structure of the IL1R–IL1 β complex.

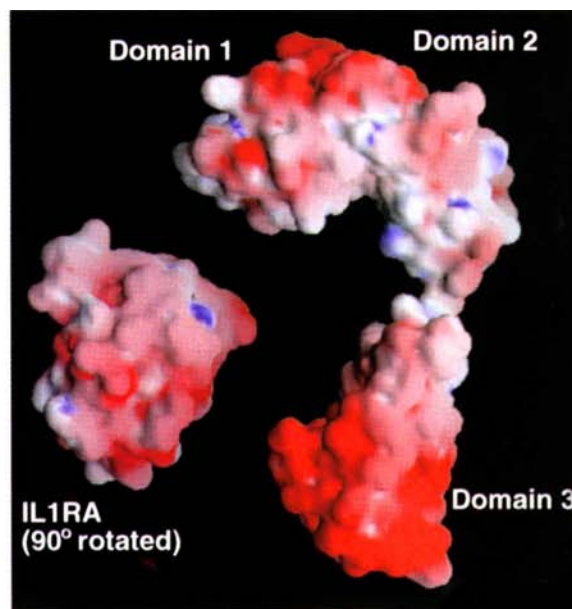


Figure 7 Solvent-accessible surface of the IL1R, coloured according to the electrostatic potential using GRASP³⁰. The contours run from $-5K_B T$ (red) to $+5K_B T$ (blue), where K_B is the Boltzmann constant and T the absolute temperature. The IL1RA molecule is peeled off and rotated by 90° to show the receptor-binding region.

domains are connected by relatively short linkers which allow flexibility while maintaining contact between the domains. Domains 1 and 2 of IL1R are close together and have tight and extensive contacts. The upper halves of strands f2 and g2 (Fig. 2) in particular have more contacts with domain 1 than with domain 2 to which they belong. The centres of mass of domains 1 and 2 are $\sim 8 \text{ \AA}$ closer in IL1R than they are in HGHR. Judging from the crystal structure, it seems unlikely that domains 1 and 2 are able to move with respect to each other, suggesting that domains 1 and 2 could behave as a single module. The connection between domains 2 and 3, on the other hand, is a long and flexible linker with no direct contacts between the domains. We would therefore expect that domain 3 is able to move with respect to domains 1 and 2 when no ligand is present.

An IL-1-receptor-like protein, the IL1R accessory protein, may be involved in the IL-1 response¹⁶. Association of this accessory protein with the complex of the IL1R with the agonists IL-1 α or IL-1 β is believed to trigger the receptor response. The accessory protein might need binding sites both on the IL1R and on the agonists IL-1 α or IL-1 β , or alternatively, the agonists may induce a conformational change in the receptor which allows binding of the accessory protein.

We investigated these possibilities by constructing a preliminary model of IL-1 β bound to the receptor, in which the structure of IL-1 β ¹⁷ was superimposed onto IL1RA in the complex by means of a procedure described before⁵. We compared this model of the complex with the extensive mutagenesis data available for IL-1 β ⁵. Some IL-1 β residues that are important for binding (Arg 11, His 30, Asn 108 and Glu 128) are all close to the receptor after superimposition and face receptor domains 1 and 2, indicating that similar regions on the surface of IL-1 β and IL1RA may interact with receptor domains 1 and 2.

The situation is quite different for residues facing domain 3. Whereas Lys 92 and Lys 94 are part of the proposed receptor trigger site², after superimposition they are facing the solvent and do not interact with the receptor. They are part of a highly positively charged region on IL-1 β involving Arg 4, Lys 92, Lys 93 and Lys 94, about 15 Å away from a region rich in negative charged on the receptor, involving Asp 248, Glu 249, Asp 251, Glu 256 and Asp 257 (Fig. 6). Electrostatic calculations (Fig. 7) indicate that this negative patch, together with a negative patch on top of domain 1, are the strongest features on the relatively featureless electrostatic surface. The close complementarity of the positive and negative sites suggests that they might interact directly with each other. A rotation of domain 3 of $\sim 20^\circ$ would bring the two sites together. The flexible linker between domains 2 and 3 allows such a rotation, which would make the conformation of the IL1 β -receptor complex more closed than that of the IL1RA-receptor complex, inducing binding of the accessory protein and hence activation of the receptor. Without a crystal structure of the IL1 β -receptor complex, however, we cannot exclude the possibility that the accessory protein will simply bind between the two charged regions.

To test the importance of domain 3 for agonist and antagonist binding, we prepared a truncated receptor consisting only of domains 1 and 2. This truncated receptor binds IL-1 α and IL-1 β with low affinity (dissociation constant (K_d) values of $7 \mu\text{M}$ and $>10 \mu\text{M}$, respectively) whereas it binds IL1RA with high affinity (K_d 28 nM). This shows that agonists, but not antagonists, need domain 3 for high-affinity binding and suggests that domain 3 strongly interacts with IL-1 agonists.

Residue 145 (Asp in IL-1 β , Lys in IL1RA), which is crucial in determining agonist or antagonist activity¹⁸, does not interact directly with the receptor, either in the IL1RA complex or in our model of the IL-1 β complex (Fig. 6), suggesting that this residue might be important for interaction with the accessory protein.

The structure presented here illustrates in detail a new binding

mode for cytokine receptors and provides insight into the possible differences between agonist and antagonist interaction. □

Methods

Equipment and crystals. The IL1R-antagonist complex was crystallized as described⁴. Native and heavy-atom data were recorded with a Siemens X1000 area detector, mounted on a Siemens rotating-anode generator equipped with a copper anode and graphite monochromator, operating at 50 kV and 90 mA. The crystals had dimensions of $\sim 0.2 \times 0.3 \times 0.5 \text{ mm}^3$ and one crystal was used per data set. Data were processed using XDS software¹⁹. The space group was $P2_12_12_1$, with $a = 47.2 \text{ \AA}$, $b = 84.6 \text{ \AA}$ and $c = 140.2 \text{ \AA}$.

Structure solution. The structure was solved with a combination of molecular replacement and heavy-atom derivatives. For molecular replacement, we used the AMoRe package²⁰ and data between 15.0 and 4.0 Å. The structure of the free IL1RA as solved by us⁵ gave clear solution with a correlation of 25.8% (R factor = 51.3%). We tried structures of various protein domains with immunoglobulin folds but most of them did not produce any clear signal, probably because these models were too different from the receptor domains and because one receptor domain represents only a small fraction of the total contents of the asymmetric unit of the crystals. A truncated model of the first CD4 domain²¹, however, gave a correlation of 29.2% in combination with the IL1RA model (R factor 50.3%). The maps, based on this partial model, were not readily interpretable and it was therefore decided to search for heavy-atom derivatives. We obtained two independent heavy-atom derivatives and one double derivative. Scaling of heavy-atom derivative data, calculation of Pattersons and refinement of heavy-atom parameters were done by using programs from the CCP4 suite²². Heavy-atom positions were located in difference Patterson maps and the parameters were refined with HEAVY²³. Phases were calculated with MLPHARE²⁴. Cross-difference Fouriers were used to define a common origin and to confirm the sites. Difference fouriers using phases from the partial model, obtained by molecular replacement, did not reproduce the sites, indicating that the model phases were rather poor. This is most probably caused by the large fraction of the diffracting matter (two receptor domains), which is missing in the partial model. However, phased translation functions²⁵ using phases from the partial model and single isomorphous replacement (SIR) phases calculated with the $\text{Hg}(\text{CN})_2$ and K_2PtCl_4 derivatives, defined a common origin and indicated the correct hand for the heavy-atom solution. The MIR map, calculated with three derivatives, was improved by solvent flattening and histogram matching using SQUASH²⁶. Map interpretation and model building were done using the program O²⁷. The model of IL1RA needed only minor rebuilding, but the model of the truncated CD4 domain needed major rebuilding to become an IL1R domain; two additional receptor domains had to be built from scratch. The structure was refined by simulated annealing and energy minimization using X-PLOR²⁸. The final model contains residues 7–151 of the IL1RA and residues 1–311 of the IL1R.

Binding studies. Truncated receptors were expressed on the surface of CHO cells and radiolabelled peptides were used for competitive binding¹¹.

Received 4 November 1996; accepted 4 February 1997.

1. Baumann, H. & Gauldie, J. The acute phase response. *Immunol. Today* **15**, 74–80 (1994).
2. Dinarello, C. A. The interleukin-1 family: 10 years of discovery. *FASEB J.* **8**, 1314–1325 (1994).
3. Evans, R. J. *et al.* Mapping receptor binding sites in interleukin (IL)-1 receptor antagonist and IL-1 β by site-directed mutagenesis. Identification of a single site in IL-1ra and two sites in IL-1 β . *J. Biol. Chem.* **270**, 11477–11483 (1995).
4. Schreuder, H. A. *et al.* Crystals of soluble interleukin-1 receptor complexed with its natural antagonist reveal a 1:1 receptor-ligand complex. *FEBS Lett.* **373**, 39–40 (1995).
5. Schreuder, H. A. *et al.* Refined crystal structure of the interleukin-1 receptor antagonist. Presence of a disulfide link and a *cis* proline. *Eur. J. Biochem.* **227**, 838–847 (1995).
6. Sims, J. E. *et al.* cDNA expression cloning of the IL-1 receptor, a member of the immunoglobulin superfamily. *Science* **241**, 585–589 (1988).
7. Harpaz, Y. & Chothia, C. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J. Mol. Biol.* **238**, 528–539 (1994).
8. Hemmings, J. M., Gernert, K. M., Richardson, J. S. & Richardson, D. C. The tyrosine corner: A feature of most greek key β -barrel proteins. *Prot. Sci.* **3**, 1927–1937 (1994).
9. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
10. Clackson, T. & Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383–386 (1995).
11. Yanofsky, S. D. *et al.* High affinity type I interleukin 1 receptor antagonists discovered by screening recombinant peptide libraries. *Proc. Natl Acad. Sci. USA* **93**, 7381–7386 (1996).
12. Banner, D. W. *et al.* Crystal structure of the soluble human 55 kd TNF receptor-human TNF complex: Implications for TNF receptor activation. *Cell* **73**, 431–445 (1993).

13. Vos, A. M., Ultsch, M. & Kossiakoff, A. A. Human growth hormone and extracellular domain of its receptor: Crystal structure of the complex. *Science* **255**, 306–312 (1992).
14. Walter, M. R. *et al.* Crystal structure of a complex between interferon- γ and its soluble high-affinity receptor. *Nature* **376**, 230–235 (1995).
15. Livnah, O. *et al.* Functional mimicry of a protein hormone by a peptide agonist: The EPO receptor complex at 2.8 Å. *Science* **273**, 464–471 (1996).
16. Greenfeder, S. A. *et al.* Molecular cloning and characterization of a second subunit of the interleukin 1 receptor complex. *J. Biol. Chem.* **270**, 13757–13765 (1995).
17. Priestle, J. P., Schär, H.-P. & Grütter, M. G. Crystal structure of the cytokine interleukin-1 β . *EMBO J.* **7**, 339–343 (1988).
18. Ju, G. *et al.* Conversion of the interleukin 1 receptor antagonist into an agonist by site-specific mutagenesis. *Proc. Natl Acad. Sci. USA* **88**, 2658–2662 (1991).
19. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
20. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
21. Wang, J. *et al.* Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature* **348**, 411–418 (1990).
22. Collaborative Computational Project Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
23. Terwilliger, T. C. & Eisenberg, D. Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions. *Acta Crystallogr. A* **39**, 813–817 (1983).
24. Otwinowsky, W. Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering* Proc. CCP4 Study Weekend, 25–26 January 1991 (compiled by Wolf, W., Evans, P. R. & Leslie, A. G. W.) 80–86 (1991).
25. Read, R. J. & Schierbeek, A. J. A phased translation function. *J. Appl. Crystallogr.* **13**, 490–495 (1988).
26. Zhang, K. Y. J. SQUASH—combining constraints for macromolecular phase refinement and extension. *Acta Crystallogr. D* **49**, 213–222 (1993).
27. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
28. Brünger, A. X-PLOR, version 3.1. A system for X-ray crystallography and NMR. (Yale University Press, New Haven, Connecticut, 1992).
29. Kraulis, P. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
30. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Acknowledgements. We thank J. A. Malikayil and T. Pelton for discussion and A. Bateman for advice on the classification of domains 1 and 2.

Correspondence and requests for materials should be addressed to H.A.S. (e-mail: hschreuder@frahm.hoechst.com). Coordinates will be deposited with the Brookhaven Protein Data Bank and will be released after one year.

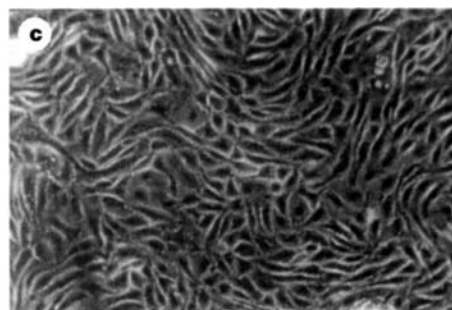
ERRATUM

Viable offspring derived from fetal and adult mammalian cells

I. Wilmut, A. E. Schnieke, J. McWhir, A. J. Kind & K. H. S. Campbell

Nature **385**, 810–813 (1997).

In this Letter in the 27 February issue, a production error led to the image for part b of Fig. 1 (fetal fibroblasts) being used twice, as parts b and c. The correct image for Fig. 1c (mammary-derived cells) is shown below, and is also on the *Nature* web site and in reprints.



YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

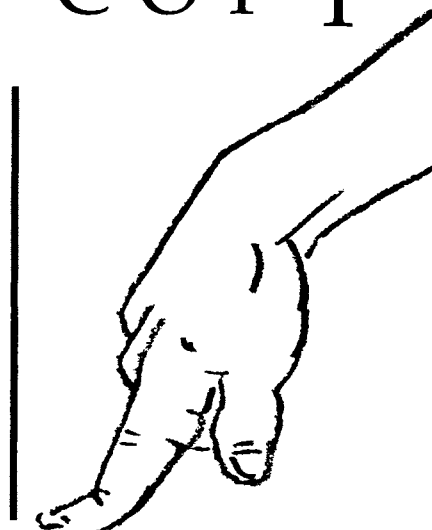
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised infringing photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986